DOCUMENT RESUME

ED 147 359 TM 006 811

AUTHOR	Dyer, Robert; And Others
TITLE	Questionnaire Construction Manual. Annex: Literature Survey and Bibliography.
INSTITUTION	Army Research Inst. for the Behavioral and Social Sciences, Ft. Hood, Tex.
REPORT NO	ARI-P-77-2
PUB DATE	Jul 76
CONTRACT	DAHC-19-74-C-0032
NOTE	465p.; For related document, see TM 006 810
EDRS PRICE	MF-$0.83 HC-$24.77 Plus Postage.
DESCRIPTORS	*Bibliographies; Internal Scaling; Interviews; *Literature Reviews; *Questionnaires; Response Style (Tests); Surveys; *Test Construction; Testing; Test Interpretation; Test Items

ABSTRACT

A literature review and bibliography on questionnaire construction are presented. The broad definition of questionnaire includes scales, structured interview forms, survey forms, and similar paper and pencil instruments used to elicit responses and collect information. A comprehensive literature search of journal articles, books, and reports in the fields of psychology, education, sociology, marketing, and the military was conducted as the basis for both this document and the related manual on questionnaire construction. Manual and computer searches of the literature included the following information retrieval systems: Psychological Abstracts, Educational Resources Information Center, National Technical Information Service, Defense Documentation Center, Bureau of the Census, Annual Reviews of Psychology, Journal of Marketing, Journal of Advertising Research, Business Periodicals Index, Public Administration Information Service, and several other bibliographies. The literature review discusses the advantages and disadvantages of various types of questionnaires, item selection, comparison of scaling techniques, effects of various item presentations, test response alternatives, perceived attitudes toward commonly used phrases, physical questionnaire characteristics, administration, respondent characteristics, interpretation of results, and suggestions for further research. Over 500 references, with descriptive keywords, are included in the bibliography. (Author/GDC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED147359

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AD

P-77-2

QUESTIONNAIRE CONSTRUCTION MANUAL

ANNEX

LITERATURE SURVEY AND BIBLIOGRAPHY

FORT HOOD FIELD UNIT

TM006 811

U. S. Army
Research Institute for the Behavioral and Social Sciences

July 1976

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

J. E. UHLANER
Technical Director

W. C. MAUS
COL, GS
Commander

Research accomplished under contract
to the Department of the Army

Operations Research Associates

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-P, 1300 Wilson Boulevard, Arlington, Virginia 22209.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
P-77-2

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
QUESTIONNAIRE CONSTRUCTION MANUAL ANNEX: LITERATURE SURVEY AND BIBLIOGRAPHY

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Robert Dyer, J. J. Matthews, J. F. Stulac, C. E. Wright, and Kenneth Yudowitch

8. CONTRACT OR GRANT NUMBER(s)
DAHC 19-74-C-0032

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Operations Research Associates
Palo Alto, California

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
2Q763731A775

11. CONTROLLING OFFICE NAME AND ADDRESS
TRADOC Combined Arms Test Activity
Fort Hood, Texas 76544

12. REPORT DATE
July 1976

13. NUMBER OF PAGES
426

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
U.S. Army Research Institute for the Behavioral and Social Sciences
ARI Field Unit-Fort Hood HQ TCATA (PERI-OH)
Fort Hood, Texas 76544

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
Contracting Officer's Technical Representative was George M. Gividen, Chief, ARI Field Office at Fort Hood, Texas. Companion volume is "Questionnaire Construction Manual."

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Questionnaire construction
Test development
Item development
Questionnaire administration
Bibliography of test construction methods

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
A literature survey on questionnaire construction encompassed journals, books, and reports in psychology, sociology, education, and marketing, and documentation published by the Defense Department. The search yielded over 2,000 references, which are listed in the 279-page annotated bibliography. A synthesis of the findings is based on abstracts selected for their relevance from the references. Results of the search were the basis for developing a manual on questionnaire construction designed for Army personnel responsible for field evaluations.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Technical Report

QUESTIONNAIRE CONSTRUCTION

LITERATURE SURVEY

Robert Dyer, Josephine J. Matthews,
Joseph F. Stulac, Calvin E. Wright, and
Kenneth Yudowitch
Operations Research Associates

ARI FIELD UNIT, FORT HOOD, TEXAS
George M. Gividen, Chief

Submitted by:
Joseph Zeidner, Director
ORGANIZATIONS & SYSTEMS LABORATORY

Approved by:
J. E. Uhlaner
TECHNICAL DIRECTOR

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

Office, Deputy Chief of Staff for Personnel
Department of the Army
1300 Wilson Boulevard, Arlington, Virginia 22209

June 1975

Army Project Number
2Q763743A775

Contract Number DAHC19-74-C-0032

FOREWORD

There has been a continuing concern within the Army Research Institute for the Behavioral and Social Sciences (ARI) for the improvement of our procedures of obtaining information, judgments, evaluations from people through such instruments as the questionnaire, survey, and interview. This concern cuts across all the technical areas of ARI and is shared by many researchers throughout the services. In FY 74 the ARI Field Unit at Modern Army Selected Systems Test, Evaluation and Review (MASSTER) at Fort Hood, Texas contracted with Operations Research Associates of Palo Alto, California to perform three tasks related to questionnaires: one, to survey the literature and abstract articles relevant to the construction and administration of questionnaires or which identified and assessed the influence of factors on responses to questions; two, to conduct experimentation on gaps that the literature survey showed to exist in the "data base", three, to prepare an instructional manual on the construction and administration of questionnaires and structured interviews.

Mr. George M. Gividen, Chief of the ARI Fort Hood Field Unit, was COTR for this contract. The contractor's effort was directed by Dr. Kenneth L. Yudowitch. In the FY 77 ARI Advanced Development Program, this effort falls under Program Element 6.37.43.A, under the project titled "Human Performance in Field Assessment" (A775).

QUESTIONNAIRE CONSTRUCTION LITERATURE SURVEY

BRIEF

Requirement:

To survey and review the literature on the design and construction of questionnaires, seeking to synthesize structural and procedural improvements and to identify gaps in our knowledge of the effects on responses to questions of various factors.

Procedure:

The literature search encompassed journals, books, and reports in the fields of psychology, sociology, education, marketing, and documentation published by the Department of Defense. Both "hand" and computer searches were made.

Findings:

The literature search yielded over 2,000 references on questionnaire methods. Abstracts were available or prepared by the contractor for about 1,000 articles. The findings are organized in twelve chapters and are followed by a 279 page bibliography.

Utilization:

The content of this literature survey may be of greatest interest and value to full-time or professional researchers who employ questionnaires in their researches. However, a significant portion of the content has entered into the writing of an ARI instructional manual on the construction of questionnaires. The manual was prepared for use by personnel charged with the development of questionnaires for use in Army field tests and evaluations.

TABLE OF CONTENTS

Chapter Page

I INTRODUCTION I-1

II ADVANTAGES AND DISADVANTAGES OF VARIOUS TYPES OF QUESTIONNAIRES II-1
  Methods to Measure Attributes and Behavior II-1
  Comparison of the Structured Interview and Mail Questionnaires II-1
  Comparison of the Structured Interview and Other Questionnaires II-3
  Comparison of Open- and Closed-Ended Items II-4
  Conclusions II-5

III SELECTION OF QUESTIONNAIRE ITEMS III-1
  Content of Questionnaire Items III-1
    Methods for Determining Questionnaire Content III-1
    Other Considerations Related to Questionnaire Content III-2
  Pros and Cons of Various Types of Questionnaire Items III-3
    Ranking Items III-3
    Rating Scale Items III-5
    Multiple Choice Items III-9
    Forced Choice and Paired Comparison Items III-11
    Card Sorts III-14
    Semantic Differential Items III-15
    Other Types of Items III-17
    Conclusions Regarding the Pros and Cons of Various Types of Questionnaire Items III-19

IV COMPARISON OF SCALING TECHNIQUES IV-1

V EFFECTS OF VARIATION IN PRESENTATION OF QUESTIONNAIRE ITEMS V-1
  Mode of Items V-1
  Wording of Items V-1
  Clarity of Items V-13
  Difficulty of Items V-15
  Length of Question Stem V-18
  Order of Question Stems V-19
  Order of Response Alternatives V-25

VI NUMBER OF RESPONSE ALTERNATIVES AND RESPONSE ANCHORING VI-1
  Issues Regarding Number of Response Alternatives to Employ VI-1
  Response Anchoring VI-9

VII ORDER OF PERCEIVED FAVORABLENESS OF COMMONLY USED WORDS AND PHRASES VII-1
  Major Studies and Lists of Adjectives and Scale Values VII-1
  Summary and Conclusions VII-29

TABLE OF CONTENTS (Cont.)

Chapter — Page

VIII CONSIDERATIONS RELATED TO THE PHYSICAL CHARACTERISTICS OF QUESTIONNAIRES VIII-1
- Location of Response Alternatives Relative to Stem VIII-1
- Questionnaire Length VIII-1
- Questionnaire Format Considerations VIII-2
- The Use of Answer Sheets VIII-3

IX CONSIDERATIONS RELATED TO THE ADMINISTRATION OF QUESTIONNAIRES IX-1
- Effects of Instructions IX-1
- Effects of Various Motivational Factors IX-2
- Effects of Anonymity IX-6
- Effects of Administration Time IX-9
- Effects of Characteristics of Questionnaire Administrators IX-10
- Effects of Administration Conditions IX-13
- Effects of Other Factors Related to Questionnaire Administration IX-14

X CHARACTERISTICS OF RESPONDENTS THAT INFLUENCE QUESTIONNAIRE RESULTS X-1
- Item Format Biases X-1
- Social Desirability Response Set X-2
- Acquiescence Response Set X-3
- Extreme Response Set X-4
- Effects of Attitudes on Responses X-5
- Effects of Demographic Characteristics on Responses X-6
- Summary and Conclusions X-7

XI CONSIDERATIONS RELATED TO THE EVALUATION OF QUESTIONNAIRE RESULTS XI-1
- Scoring of Questionnaire Results XI-1
- Properties and Uses of Ipsative Scores XI-3
- Data Analyses XI-6

XII RECOMMENDED AREAS FOR FURTHER RESEARCH XII-1
- Advantages and Disadvantages of Various Types of Questionnaires XII-1
- Selection of Questionnaire Items to be Used XII-1
- Comparison of Scaling Techniques XII-2
- Effects of Variation in Presentation of Questionnaire Items XII-2
- Number of Response Alternatives and Response Anchoring XII-3
- Order of Perceived Variables of Commonly Used Words and Phrases XII-3
- Considerations Related to the Physical Characteristics of Questionnaires XII-3
- Considerations Related to the Administration of Questionnaires XII-3
- Characteristics of Respondents that Influence Questionnaire Results XII-4
- Considerations Related to the Evaluation of Questionnaire Results XII-4
- General Recommendations XII-4

BIBLIOGRAPHY B-1

LIST OF TABLES

Page

Table V-1: Summary of Studies on Mode of Items V-2
Table V-2: Summary of Research on Positive versus Negative Wording of Items V-5
Table V-3: Summary of Research on Objective versus Subjective Wording of Items V-9
Table V-4: Summary of Literature on Item Difficulty V-16
Table V-5: Summary of Studies Relating to the Order of Question Stems V-20
Table V-6: Summary of Studies on Order of Response Alternatives V-27

Table VI-1: Summary of Studies Relating to Number of Response Alternatives VI-2

Table VII-1: Scale Values of Standard Set of Words, VII-2

Table VII-2: Scale Values of Selected Words VII-2
Table VII-3: Words Marked "Unable to Rate" by 20 or More Subjects VII-3
Table VII-4: Words Exhibiting Marked Bimodality of Response VII-3
Table VII-5: Scale Values as Affected by Adverbial Modifiers VII-3
Table VII-6: Scale Values and Standard Deviations of Stimulus Items VII-5
Table VII-7: Means and Standard Deviations of Commonly Used Statements VII-6
Table VII-8: Obtained Successive Intervals Scale Values of Adverb-Adjective Combinations VII-7
Table VII-9: Adverb and Adjective Value Matrices VII-8
Table VII-10: Numerical Ratings of Adverb-Verb Combinations VII-9
Table VII-11: Scale Positions for Thirty-four Phrases VII-10
Table VII-12: Scale Positions of 47 Intensity Phrases VII-11
Table VII-13: Stability of Intensity Phrases in Diverse Contexts VII-12
Table VII-14: Ratings of Likableness, and Likableness Variances for Personality Traits VII-15
Table VII-15: Means and Standard Deviations for Phrases of Degrees of Adequacy VII-20
Table VII-16: Means and Standard Deviations for Phrases of Degrees of Acceptability VII-22
Table VII-17: Means and Standard Deviations for Phrases Used for Comparison VII-24
Table VII-18: Scale Scores of Statements Based on Over-all Acceptability VII-25
Table VII-19: Meaning of Frequency of Words VII-26
Table VII-20: Correlations of Jones & Thurstone and Myers & Warner "Scale" Values VII-27
Table VII-21: Correlations of Myers-Warner and Cliff Scale Values VII-27
Table VII-22: Summary of Perceived Favorableness of Commonly Used Words and Phrases VII-28

Chapter I

INTRODUCTION

ORA-Operations Research Associates has surveyed and reviewed the literature on the design and construction of questionnaires as part of a contract with the Army Research Institute for the Behavioral and Social Sciences, Fort Hood, Texas. This report presents the results of that survey and review. It is based on a broad definition of questionnaire to include scales, structured interview forms, survey forms, and similar paper and pencil instruments used to elicit responses and collect information.

The emphasis of this review was on questionnaires used with Army personnel participating in military field tests concerned with evaluating training, equipment, organizations, concepts, and doctrine, but little was found on this topic. However, since considerations affecting questionnaire construction for Army field test evaluations are common to questionnaire construction for other uses, this review covers the pertinent literature from other fields. The review was not concerned with the evaluation of soldier attitudes or reactions pertaining to societal problems, personality, academic testing, or similar research areas except as the related methodological considerations were also applicable to field testing. Emphasis was placed on those sources which provided empirical data on questionnaire construction. Material on the administration and analysis of questionnaires and on questionnaire application and results was excluded except where specifically related to questionnaire construction. Topics not stressed in the literature review are noted as appropriate in the text.

The literature search was quite comprehensive and included the review of journals, books, and reports in the fields of psychology, education, sociology, marketing, and the military. Both hand and computer searches were made. Computer searches were made of information retrieval systems maintained by: the American Psychological Association for Psychological Abstracts covering the years 1967 to 1974; the Educational Resources Information Center for the years 1957 to 1974; the National Technical Information Service for 1963 to 1974; the Defense Documentation Center; and the Bureau of the Census.

Hand searches were made to supplement the computer searches and included the:

Psychological Abstracts for 1949 through 1967; Annual Reviews of Psychology for 1960 through 1974; Journal of Marketing for 1942 to 1974; Journal of Advertising Research for 1960 to 1974; Journal of Marketing Research for 1964 to 1974; Business Periodicals Index for 1951 to 1974; and Public Administration Information Service for 1949 to 1974.

Hand searches were also made of several bibliographies: Goheen and Kavruck (1950) covered the early work for the years 1929 to 1949; Potter, Sharpe, Hendee, and Clarke (1972) covered more recent work; the ARI Field Unit at MASSTER, Fort Hood, provided a March, 1974, short bibliography on the subject; and the in-process bibliography of the Army's Test and Evaluation Command was also reviewed. Finally, the articles abstracted were reviewed for references, as were recognized pertinent texts and staff personal files.

The literature search yielded a total of over 2,000 citations on questionnaire construction and methodology; however, abstracts were only available or prepared for about half of the citations. This limitation was imposed by the level of effort available, and the selections were made on the basis of the apparent relevance of each citation, judging primarily from its title or abstract if available. The actual writing of the following chapters was based on a selection from these abstracts, with occasional reference to the actual articles, depending on the organizational needs of the chapter as seen by its author. The articles actually cited in the writing are included in the attached bibliography and are identified by asterisks.

The results of this literature search were used as a basis for the development of a manual on questionnaire construction (Dyer, Matthews, Wright, & Yudowitch, 1975). The manual was prepared for use as a guide by personnel charged with the development of questionnaires for use in Army field test evaluations. It includes chapters on topics discussed in this report.

In the text which follows, Chapters II through XI were selected and organized to cover comprehensively and with minimal overlap the technical objectives of the study contract between ORA and ARI. These chapters also include for completeness some additional parallel items. Chapter II discusses the advantages and disadvantages of various types of questionnaires. Chapter III considers the selection of questionnaire items to be used, including the content of questionnaire items and the pros and cons of using various types of questionnaire items. Chapter IV notes articles about various scaling techniques. The effects of variations in the presentation of questionnaire items are covered in Chapter V, while Chapter VI reviews articles on the number of response alternatives and response anchoring. The order of perceived favorableness of commonly used words and phrases is the topic of Chapter VII. Chapter VIII examines considerations related to the physical characteristics of questionnaires, while considerations related to the administration of questionnaires are covered in Chapter IX. Characteristics of respondents that influence questionnaire results, including various biases and response sets, are discussed in Chapter X, while Chapter XI is devoted to considerations related to the evaluation of questionnaire results. Finally, Chapter XII notes recommended areas for further research based upon either identified gaps in the empirical research or contradictions among studies.

Chapter II

ADVANTAGES AND DISADVANTAGES OF VARIOUS TYPES OF QUESTIONNAIRES

This chapter discusses, to the extent articles were available on the topic, some of the advantages and disadvantages of using various types of questionnaires, as the word "questionnaire" was defined in Chapter I. In the first section below, methods to measure attributes and behavior are mentioned. Next, the structured interview is first compared with mail questionnaires, and then with other types of questionnaires. Comparisons between open- and closed-ended items are then discussed.

Methods to Measure Attributes and Behavior

There are a number of techniques of data collection that can be used to measure human attributes and behavior, some of which have been reviewed by Deri, Dinnerstein, Harding, and Pepitone (1948). The methods include observation, personal and public records, specific performances, sociometry, interviews, questionnaires, rating scales, pictorial techniques, projective techniques, achievement testing, and psychological testing, among others. For this review, however, attention has been restricted to a more limited number of data collection techniques: certain paper and pencil types of instruments broadly classed as questionnaires as defined in Chapter I, and including only some of the techniques mentioned above. A distinction has also been made, in the text to follow, between open-ended questionnaire items and closed-ended items. Open-ended items are those which permit the respondent to express his opinions in his own words and to indicate any qualifications he wishes. Closed-ended items, on the other hand, utilize response alternatives, such as multiple choice or true-false. Structured interviews are included within the definition of questionnaire used since typically an interview schedule is developed and employed by an interviewer both for asking questions and recording responses much like a self-administered questionnaire with open-ended items. This distinction is not as clear as it might be, however, since some investigators (such as Paradise and Blankenship, 1951) admit of orally administered questionnaires, structured interviews, and unstructured interviews. In any case, unstructured interviews are outside the scope of this review, and they will not be discussed further.

Comparison of the Structured Interview and Mail Questionnaires

During the literature review, attention was given to articles on the use of mail questionnaires only to the extent that the information might be generalizable to other types of questionnaires. Accordingly, any articles related to sampling considerations, correcting variance estimates for non-response, etc., were ignored. Since the use of mail questionnaires involves the consideration of issues that do not pertain to the use of other types of questionnaires, they are discussed separately.

A number of criteria was employed by O'Dell (1962), who compared personal interviews and mail questionnaires using identical terms. He found an interview bias, in that during the interviews the usage of certain types of products was understated when it might reflect unfavorably on the respondent. Wiseman (1972), compared a mailed questionnaire, telephone interview, and personal interview, concluded that issues involving socially accepted or rejected answers will effect more bias in interviews than in questionnaires. Ellis (1948) similarly found more self-revelatory or unfavorable responses in anonymous mailed questionnaires than in interviews. Ford (1969) asked identical questions in a mail questionnaire followed by an interview. He fourd that there was a consistency of response about newspaper readership and about socioeconomic factors, but inconsistency on items related to attitudes and opinions, the location of past purchases, and when past purchases were made. A number of factors may have influenced his results, however, such as the time lapse between the questionnaire completion and the interview. Williams (1968) noted that data gathered by telephone interview may be less accurate than those obtained from a mail questionnaire, since the group who are at home to answer the telephone may not be as representative as those to whom the questionnaires are mailed.

The comparative costs of interviews and mailed questionnaires were discussed in five articles. Cahalan (1951) administered a 23 page mailed questionnaire to 1,051 Army officers, and found it was less expensive, more anonymous, and faster than the interview technique. O'Dell (1962) reported that the costs of interviewer time tended to outweigh the costs of obtaining and maintaining a mail panel. Gibson and Hawkins (1968) concluded that, under the promise of anonymity, the questionnaire should equal the interview in response information at a much smaller expense (although there is some question about the survey design they employed). The degree of consistency between interview and questionnaire results found by Parker, Wright, and Clark (1957) also raised questions concerning the justification of the expense of interviewing, when questionnaires or similar techniques would be only slightly less reliable. Sudman, Greeley, and Pinto (1965) were somewhat more conservative in their conclusion, reporting that costs were not significantly affected, regardless of whether interviews, mail questionnaires, or a combination of both were employed.

Specificity of responses was discussed only by O'Dell (1962). He noted that noncommittal responses and the tendency not to answer open-ended questions were more prevalent for mail questionnaires than for personal interviews, as might be expected.

Combinations of survey methods were discussed in three articles. Sudman, Greeley, and Pinto (1965) found that self-administered questionnaires used in conjunction with personal interviews elicited a slightly higher cooperation/return rate from respondents than either used alone. The result that comparisons between interviews, self-administered questionnaires, and a combination of both did not indicate any large differences suggested to them that additional flexibility should be considered in methods of survey research. Payne (1964) also suggested that sometimes a combination of survey methods, such as personal interview, telephone interview, and mail questionnaires, may be used with the same respondents to produce results more efficiently than one method alone could do. However, he presented no firm evidence of higher reliability or validity for combined methods over individual survey methods. Sharp (1955) found that when respondents were unable to give complete information during an interview and copies of a questionnaire were left to be mailed back, 40% were returned thus eliminating the necessity for call-backs.

Comparison of the Structured Interview and Other Questionnaires

Most of the studies comparing the structured interview with questionnaires other than mail questionnaires did so in terms of the consistency of response from one technique to the other. For example, Bennett, Alpert, and Goldstein (1954), though working with only 16 subjects, found that 26 out of 30 questions showed significant consistency of response from a one hour interview immediately followed by the use of a limited response questionnaire on the same topic. Consistency coefficients reported were 1.00 (perfect) for sociological information, .78 on knowledge, .69 on past behavior, and .46 on attitudes. The conclusion reached was that on information other than sociological, differences in response will be noted between interview and limited choice questionnaires, especially concerning attitudes.

The results obtained by Bennett, Alpert, and Goldstein (1954) appear to have been supported in part by two other investigations. Walsh (1967) compared the accuracy of the interview, questionnaire, and personal data blank for collecting verifiable biographic information. Comparing collected data to available records for 270 students, he found no differences, and concluded that biographic data may be collected reliably by the most efficient means. Boulger (1970) also found that the validity of response to interviews and questionnaires was not significantly different in the elicitation of life history data.

Three studies compared structured interviews and other questionnaires in the measurement of attitudes. In the first, Metzner and Mann (1952) followed a fixed alternative questionnaire administered to 328 employees with an open-ended interview. They noted a tendency for the employees to rate slightly higher in the interview than on the questionnaire. There were, however, a number of limitations to the study, including a two month time lapse between completion of the questionnaire and the interview. In the second study, Wedell and Smith (1951) found that interviewers overestimated attitude in comparison with self-judged attitude, although the objective rating of interview record sheets was closer to self-rating than the interviewers' rating. Wheatley (1973), however, found no significant differences between mean scale scores for two groups, one of which expressed their attitudes during a telephone interview, while the other group responded on a self-administered questionnaire.

Although studies involving the use of questionnaires for the measurement of personality were generally excluded from the literature review, three were considered in that they compared results obtained from interviews and questionnaires. Eysenck and Eysenck (1962) sought to answer the question of whether an interview-questionnaire would reveal a factorial structure essentially identical to that found with questionnaires administered in the orthodox manner. The results indicated that the method of administration did not affect the factorial composition of the items, which measured extraversion and neuroticism. Ambler, Blair, deRivera, Nelson, and Schoenberger (1958) also found that the interview and questionnaire methods gave similar results in the classification of subjects according to three levels of anxiety towards flying. The conclusions reached by Levonian (1963), however, were different. He determined the reliability of three short personality scales administered by the interview survey method to 432 subjects. The values were sufficiently less than the consistency reliabilities of short scale personality measures obtained by the usual questionnaire survey method to raise serious questions about the adequacy of such personality measures obtained by the interview method.

A comparison of interview and other questionnaire results when ego-involving questions were asked was the topic of two reports. Knudsen, Pope, and Irish (1967) concluded that interviews may lessen the expression of deviance, compared with anonymous questionnaires. Based on three different samples of white women all of whom were or had been premaritally pregnant for the first time, the data suggested that in interview situations respondents were more likely to support the public and restrictive sexual norms that they assumed were adhered to by the interviewer. In the private and anonymous questionnaire situation, the respondents more often answered to subcultural norms. Ellis (1947b) compared the questionnaire and interview methods in the study of human love relationships. His results indicated that the great majority of subjects gave less favorable, or more incriminating, responses to the questionnaires than they did to the interview. Ellis concluded that for more ego-involving questions the questionnaire may produce more self-revelatory data than the interview.

Comparison of Open- and Closed-Ended Items

Of the five articles that compared the use of open-ended and closed-ended questionnaire items, three appeared to favor the use of the open-ended format, at least for the factors considered. Ellenbogen and Danley (1962), in a study of the comparability of responses to a socially concordant question, found that responses were more varied to the open-ended question than to the closed, although the closed had an "other" category. Asking about resources of helpful health advice, they also found that 19% of the responses were inconsistent, in that sources of advice cited in the open question were omitted in the closed.

England (1948) compared open-ended and dichotomous items about capital punishment in three survey samples of 2,000, 3,000, and 6,000. The results gave preference to the open-ended items, since they allowed for the expression of middle party opinions that the dichotomous items forbid. However, in coding the open-ended items, expert analysts were required to obtain reliable results.

The results of a computer-assisted method of free response (after which the respondents evaluated the responses they generated on a rating scale) was compared with responses to prelisted statements in a study by Kohan, deMille, and Myers (1972). Although no significance tests were reported, the free response method appeared to generate response categories that differed rather substantially from the prelisted statements. Issues of importance that were overlooked by the questionnaire developers were identified. It was concluded that reliance on the conventional method may distort a study's focus by obtaining data on items not of real concern and having no accurate means to measure concern. The authors also noted that high affirmative levels for an item can often be interpreted as a response set or lip service, while responses generated by unstructured methods are probably more reflective of personal involvement or concern.

The study favoring the use of close-ended items was by Scates and Yoemans (1950a). It was undertaken by the American Council on Education to determine the value of objective tests for identifying those scientists and engineers who were likely to undertake further education. It was concluded that the use of objective tests was more advantageous than the several depth essay questions used in a previous study, because they took less time and were therefore more acceptable to the employees.

The best summary for this section was stated by Prien, Otis, Campbell & Saleh (1964). They noted that the open-ended type of questionnaire has the advantage of providing unique information, whereas the objective type of questionnaire is generally more reliable. The combination of both, they said, would appear to be best.

Conclusions

The decision about which type of questionnaire to use depends upon the specific research question that one is attempting to answer and the practical limitations involved. Both structured interviews and other types of questionnaires appear to have their place in research studies, and both have have their limitations. The choice of which to use may well depend upon costs, which are generally lower for the typical questionnaire. The typical questionnaire is apparently more reliable, while the structured interview may provide more unique information. If the dimensions of a problem have not been explored before, the best compromise would appear to be to use the interview approach with open-ended items to uncover the dimensions, and follow this by the use of the more reliable paper and pencil questionnaire to obtain more specific information.

Chapter III

SELECTION OF QUESTIONNAIRE ITEMS TO BE USED

Once a decision has been made as to the type of questionnaire instrument to use (the topic of Chapter II), the specific questionnaire items to be administered need to be selected. The two main sections in Chapter III, then, address the content of questionnaire items and the pros and cons of various types of questionnaire items.

Content of Questionnaire Items

This section considers first methods for determining questionnaire content, and then other issues related to questionnaire content.

Methods for Determining Questionnaire Content

There are a number of ways that can be used to determine questionnaire content. One of these that is not too well known is the critical incident technique. As noted by Flanagan (1954) the critical incident technique consists of a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in both solving practical problems and in developing broad psychological principles. The technique outlines procedures for collecting observed incidents of behavior having special significance and meeting systematically defined criteria. It can be of assistance, therefore, in helping to determine the content of items to be included in questionnaires. Although many articles on the technique have been published, they were not all reviewed in conjunction with preparing this review. One article on the topic was prepared by Barnes (1960), who gave an historical sketch of the development of the technique,and outlined the procedures to follow in using this approach for social research. The procedures, representing one way that the critical incident technique can be used, included: determining the aims of the investigation; securing competent reporters or observers; collecting the critical incidents of behavior actually observed; selecting those incidents to be included in the final study; analyzing and classifying the data; and interpreting the findings.

Another method for selecting items for an attitude scale was used by Alilunas (1949), who was concerned not only with finding out what people think about an issue, but how they think about matters on which they are asked to give an opinion. The method starts with asking a group of individuals to write six statements giving their impressions of a topic, such as capitalism. From these, some smaller number of statements are selected that are readable, intelligible, and capable of classification. These statements can then be sorted into several categories, such as the status of the topic and its good and bad features.

Yet another way of developing closed-ended questionnaire items is to evaluate the responses to corresponding open-ended items, as suggested by authors such as Payne (1965). Reporting on a computer-assisted method of free response analysis where respondents give answers they think appropriate and then rate their answers according to dimensions specified on a scale, Kohan, deMille, and Myers (1972) stated that the method identified issues of importance that had been overlooked by questionnaires developers either because of their own biases or imperfect knowledge. They also noted that reliance on the conventional method may distort a study's focus by obtaining data on items not of real concern and having no accurate means to measure concern. Also on the topic of bias, Schuessler (1952) questioned the randomness of item selection in scale analysis. He showed data to indicate that differences among investigators' definition of the universe and bias in selecting items both effect their results. He concluded that much more critical attention is needed by the researchers to avoid their own biases and influence in gathering data for analyses.

Hart, Faust, Rowland, and Lucier (1964), in a report on attitudes of troops in the tropics, noted that the sentence completion technique is useful for assessing topic and dimension saliency, and for validating the objective techniques. They also reported that a listing technique is valuable for identifying salient topic dimensions and salient topics, and for updating instruments which are developed on pilot samples and used on larger populations. They feel that considerable effort should be exerted to identify the salient topical dimensions, their levels, and their interrelationships whenever an objective scaling technique is used.

Other Considerations Related to Questionnaire Content

This section discusses a number of diverse topics, all of which are related in some way to questionnaire content.

Five obstacles to the selection of appropriate questions to test social-psychological variables were discussed by Bradburn (1970). They are:

1. Lack of agreement among behavioral scientists about the appropriate social-psychological dependent variables that are relevant to particular social programs;

2. An inadequate conceptualization of those social-psychological variables that are suggested for study;

3. A relative lack of interest in systematic methodological research and survey measurement;

4. The relative underdevelopment of measurement theory in survey work; and

5. The special historical and cultural problems that affect the phraseology of questions.

Among the principles reported by Blankenship (1942) that he believes should be followed in the wording of preference-type questions, those relating to questionnaire content are: to be psychologically sound, a question should ask about past, present, or future behavior, rather than hypothetical opinion; the questions should not damage the pride of respondents; and the first few questions used must secure rapport with the respondent.

The fact that questionnaire items can produce variable distortion was pointed out in the report of a study by Klein, Maher, and Dunnington (1967). Items dealing with salary and with ratings of top management produced consistent positive distortions, whereas items dealing with work pressures and the respondent's manager produced little or no distortion even under conditions of high threat. Dunnette and Heneman (1956) also noted that a threat to anonymity results in differential amounts of response distortion, depending upon the content of different items comprising the questionnaire.

Marquis, Marshall, and Oskamp (undated) reported on a study of the accuracy and completeness of testimony as a function of kind of questions. They found that for items of low salience, structured questioning resulted in more complete but less accurate responses. However, for items of high salience, more structured questioning did not reduce accuracy. Similarly, Miklich (1966) found that, if an ambiguous item was important, the tendency was to agree with it. If it was unimportant, the tendency was to disagree.

Two studies considered the reliability of various types of questionnaire items as a function of content. Cavan (1933) concluded that questions involving attitudes or estimates have lower reliability than factual questions, and that reliability is increased by avoiding fine detail. Guber and Gerberich (1946), on the other hand, found that factual questions showed the least reliability.

Finally, Spector (1957) demonstrated that the test user's values and needs do, and should, enter into judgments made during the construction and validation of an attitude test.

Pros and Cons of Various Types of Questionnaire Items

This section presents the pros and cons of various types of questionnaire items as obtained from the literature reviewed. Included are: ranking items; rating scale items; multiple choice items; forced choice and paired comparison items; card sorts; semantic differential items; and other types of items. As appropriate, comparisons of item types are included, except for a comparison between open-ended and closed-ended items generally, which was discussed in Chapter II.

Ranking Items

Comparison of ranking and rating scales. Five articles were abstracted that compared ranking and rating methods. Bittner & Rundquist (1950) described th rank comparison rating method,and noted that comparisons with other studies revealed that the method gives results closely related to rank comparison. Murphy, Bailey, and Covell (1954) found, in judging frozen strawberries, that rating provided better discrimination than ranking when ten judges were used. However, Rennick, Grupe, Reich, and Sewell (1954) found rankings to be more reliable than ratings when professional staff both ranked and rated parents' descriptive reports of their children's growth in specific character attitudes. Bartlett, Heermann, and Rettig (1960) found that, for a single judge, the ranking and paired comparison techniques were superior in reliability to the Likert, graphic rating, and equal appearing intervals techniques. Kassarjian and Nakanishi (1967), however, found comparability between ranking and Likert-type scaling based on reliabilities and inter-method correlations when methods were compared for the selection of a brand name for a ficticious new phonograph.

Comparison of ranking and paired comparisons. There appears to have been contradictory evidence obtained when the ranking and paired comparisons methods were compared. Wilkins (1950), using 300 men randomly selected from British army reception centers, found that the two methods did not yield similar results when the importance of eight characteristics of jobs were considered. The observed differences did not appear to be systematic or biased, although characteristics of least importance varied the most between methods. Witroyl & Thompson (1953) found that a paired comparison questionnaire was a more stable measure than a partial rank order form of a social acceptance questionnaire administered to about 80 sixth grade students. They noted that this may be due to the larger number of responses required in the paired comparison form. They also said that this form is a more sensitive measure of the status of individuals in the middle range of the acceptability continuum, and offers relatively more general measures of social status. The partial rank order scales may reflect more personal and situational factors. Also in favor of paired comparisons, Cohen (1967) suggested that the ranking of stimuli produces a statistical artifact that can be corrected and controlled by paired comparison analysis. The artifact is the inability of ranking to detect the comparative position of each stimulus in relation to each other stimulus. Penner, Homant, and Rokeach (1968) compared the rank order and paired comparison methods of measuring terminal and instrumental values. For the terminal values, the paired comparison reliability was significantly higher, while for instrumental values the difference was not significant, the trend being in the opposite direction. The authors concluded, however, that the benefit of the paired comparison method as compared with the rank order method is doubtful, The results suggested that the paired comparison method should be employed when measuring value systems only if there is a principal concern with the terminal values and if the time and effort expended in testing, scoring, coding, etc., are not important considerations. Bernard (1933) had come to a similar conclusion, noting that the method of ranking is not inferior in reliability to that of paired comparisons. He also stated that, since it took twice as long for the judges to complete the paired comparisons as the ranking, the latter was the superior method.

The results from three other studies found essentially no differences between the methods of ranking and paired comparisons. Eng and French (1948), comparing sociometric and psychological methods of scaling, found a near perfect correlation between mean ranks and paired comparisons. Kassarjian and Nakanishi (1967), in the study noted in the previous section, also found comparability between ranking, paired comparisons, and open choice preferences. Slater (1965) also found, from four experiments, comparability of ranking, paired comparisons, and other forced choice comparisons for recording personal preferences. He concluded that the whole weight of the evidence is in favor of the view that an informant, when expressing his personal preferences, tends to maintain a level of reliability which characterizes him as an individual, and is unaffected either by variations in the number of objects he is given to compare or changes in the methods he is asked to use.

The relationship between ranking and the method of paired comparisons was reported by Ross (1955) and Paull (1968). Ross showed that, when N judges are asked to indicate their preferences for n items by both the method of paired comparisons and the method of rank order, a linear relationship holds between the total number of choices from the paired comparison method and the mean rank from the rank order method. Paull (1968) studied the reliability of results obtained by the psychophysical methods of rank ordering and paired comparisons when subjects are ego involved in the material being judged. He found that scales derived by the two methods are linear in relationship.

Rating Scale Items

Comparison of rating scale and multiple choice items. Only one study comparing rating scale and multiple choice items is reported here, since a majority of such studies involve issues regarding the number of response alternatives to employ and are hence discussed below in the first section of Chapter VI. Greenwald and O'Connell (1970) conducted a study to test previous findings that suggested that dichotomous measures yielded similar but not equivalent information to that of Likert scales. The results showed that, as in previous studies, the true-false and Likert methods correlated significantly. However, the Likert format produced the higher item-total correlations. The greater internal consistency for the Likert approach suggested a possible advantage for Likert scaling.

Comparison of rating scale items and forced choice or paired comparison items. In this chapter, forced choice and paired comparison items are discussed together since the latter is but a special case of the former, using duads.

A study by Pilgrim and Wood (1955) compared the sensitivity of rating scale and paired comparison methods for measuring consumer food preferences under laboratory conditions. The methods were found to be equally sensitive whether the differences in preference were large or small. Similarly, Greenberg (1963) found no significant differences between rating scale and paired comparison tests used in consumer product tests.

In the attitude measurement area, Neidt and Merrill (1951) compared five point rating scales and paired (positive and negative) statements. They found that each showed about equal validity coefficients. Although the reliability of the rating scale was somewhat higher than that of the paired comparison form, the authors felt that there are advantages to the latter which warrant its consideration under some circumstances. Horst and Wright (1959) also obtained higher reliability for a self-appraisal personality rating scale than for a paired comparison inventory composed of the same items, although the rating scale scores were arithmetically ipsatized. (See Chapter XI for a discussion on the properties and uses of ipsative scores.) The rating scale also required only about one-third the time to administer than the paired comparison test.

A personality questionnaire and forced choice personality test were compared by Gordon (1951). Both had the same factor structure and much the same item content, and were constructed by the method of internal consistency. For all four personality scales the forced choice method was found to be more valid than the questionnaire method, using descriptive nominations by associates as the criterion. Multiple correlations indicated that the questionnaire data added nothing towards the prediction of the criteria when placed in a battery with the forced choice test.

Scott (1968) did a study of the comparative validities of self-report forced choice and single stimulus tests. He noted that the generalization that forced choice personality inventories are more valid than single stimulus forms of the same tests was not supported by initial examination of the relevant evidence. Apparently only one study that claimed superior validity for the forced choice format appeared to have used identical items in the two forms. Other studies either did not use single stimulus forms for comparison, did not hold item content constant between the two forms, or else yielded nonconfirming results. He reported also that the most tenable conclusion is that test validity does not depend on this characteristic of item format under the circumstances in which these self-report inventories are typically administered.

Newhall (1954) compared the methods of paired comparison and single stimuli in the evaluation of a series of color prints and color transparencies. The two methods produced highly correlated results. However, the method of single stimuli was preferred as being the more efficient method in making judgments where items do not require juxtaposition.

All of the studies to be reviewed in the rest of this section involved the use of judges or raters, and hence may not be comparable to studies based upon self-report. Using three groups of judges, the paired comparison and equal appearing intervals methods of scaling attitude statements toward teaching were compared empirically by Crawford (1965). He said that the two methods appeared comparable, at least when expert judges were used. The use of students as judges in this type of study was questioned. Students rated occupational status in the study by Bartlett, Heermann, and Rettig (1960), where little difference in scale values or reliability for mean scale values was found using the paired comparisons, Likert, graphic rating, and equal appearing intervals techniques. The paired comparison and ranking techniques were, however, found to be superior in reliability for a single judge. Using 85 subjects to judge the esthetic value of seven handwriting specimens, Ekman and Kunnapas (1960) constructed an interval scale by the method of paired comparisons and a ratio scale by a variant of the method of ratio estimation. They both gave essentially the same results.

A graphic rating scale was compared with six kinds of forced choice forms for rating Air Force technical instructors, by Berkshire and Highland (1953). Scores from the graphic rating scale exhibited relatively little bias and had as high validity as the best of the forced choice scales. Combining the scores from the graphic and forced choice scales yielded validity coefficients substantially higher than for either alone. The use of forced choice items and both eight and five-step graphic scales was compared in a study conducted by the U.S. Department of the Army (1952) using 400 officers as a rater-ratee population. The eight-step graphic scale had the highest validity (.53). A study by Staugas and McQuitty (1950), however, found the forced choice method superior to the use of a graphic scale. But Bayroff, Haggerty, and Rundquist (1954) found that two types of graphic rating scales and two modifications of the forced choice technique did not differ markedly in validity.

Susceptibility to errors was the concern of two other investigations. Leftwich and Remmers (1962) compared graphic and forced choice (tetrad) ratings of teacher performance. Distributions and intercorrelations of mean item and mean total scores showed the graphic form relatively more susceptible to errors of leniency and halo. Item intercorrelations were also higher in general for the graphic form. The authors noted, in addition, that the forced choice form was susceptible to fakability, relative to the transparency of any forced choice tetrad. Bartlett and Sharon (1969) determined the effects of several instructional rating conditions on leniency on a graphic and forced choice rating scale. A significant leniency effect was found with the graphic ratings which were to be used for evaluation purposes and those which had to be justified to the ratee, but presumably not with those that were anonymous or were identified by having the rater place his name on the form. It was concluded that the forced choice scale was quite resistant to leniency bias.

Comparison of rating scale items and card sorts. Two studies compared the use of rating scales and card sorts, both in terms of determining scale values. Seashore and Hevner (1933) substituted a nine point scale for each item for the standard method of sorting items into nine piles from separately printed slips. The rating method saved 87% of the time in assembling materials and 50% of the time in tabulating results involved in making attitude scales by Thurstone's method of equal appearing intervals. The subjects found the task easier and more pleasant, and the results showed negligible differences in the medians or scale values of the items, and in the difference or spread of opinion (Q value) in regard to them.

An investigation of the stability of median and Q values computed from graphically derived and from sorted judgments used in scaling by the method of equal appearing intervals was conducted by Siegel and Siegel (1962). Graphic judgments using a nine point scale tended to yield higher Q values than nine pile sorts for relatively unambiguous items. The medians derived from the two procedures correlated .97.

Comparison of rating scale and semantic differential items. In the study by Hart, Faust, Rowland, and Lucier (1964) on the attitudes of troops in the tropics, it was concluded that Osgood's semantic differential technique was clearly superior to Likert's agree/disagree method of summated ratings. They went on to note that, for most purposes, attitudinal data collection efforts in which objective questionnaires are used should consist of some form of the semantic differential scaling technique as opposed to agree/disagree versions of Likert's method of summated ratings.

Hughes (1967) compared the use of Thurstone and modified semantic differential scales (with a "no information" category) in a questionnaire. None of the Thurstone scales detected attitude change, but 28% of the semantic differential scales did. Test-retest reliability was .53 for Thurstone, .58 for the semantic differential. In addition, the semantic differential increased in preference as the respondents became used to it.

Ward (1969), questionning the results of a previous study, found that the semantic differential is no more vulnerable to changes in issue saliency than are other widely used measures of attitude.

Comparison of rating scale and check list items. The study by Hughes (1967) referred to immediately above also included adaptations of a check list (e.g., with important, unimportant, and no opinion categories) on the questionnaire employed. Eleven percent of the check list scales detected attitude change, and the check list items had a test-retest reliability of .58, the same as the semantic differential items.

Likert-type scaling was compared with the use of various types of check lists by Kassarjian and Nakanishi (1967), and no differences in results were found. In the Department of Army study (1952) referred to previously, the eight-step graphic rating scale was also found to have higher validity (.53) than when a controlled check list was used (.44). The four five-step scales had validities ranging from .39 to .44.

Multiple Choice Items

As used here, multiple choice items include true-false and yes-no, and similar dichotomous items as special cases. Generally, studies having to do with right/wrong responses were excluded from the literature review, unless they appeared to have direct relevance to the use of multiple choice items in questionnaires. Comparisons of rating scale and multiple choice items were reported above.

Some issues related to the use of multiple choice items. A number of issues appeared in the literature related to the use of multiple choice items. Those not more appropriately discussed in other chapters are reviewed here.

Swordes (1952) discovered that in a test using items with both four and five choices, a number of those taking the test marked the fifth space on the answer sheet when the question had only four possible responses. It was concluded that certain preventions should be taken to reduce the undesirable results of using a different number of distractors in the same examination. These include special instructions, reduction of the number of alternate groups, and restriction of a varying number of choices to the more capable test takers, when practicable.

Mosier and Price (1945) presented a means to overcome the usual problem in multiple choice items construction: the arrangement of the response alternatives. They used a table in which the 120 permutations of the numbers one through five had been randomized. They pointed out, however, that such a table should not be used when the response alternatives form a logical pattern.

Cronbach (1941a) compared multiple choice and multiple true-false tests, with instructions to guess. He found little significant differences between them. However, the multiple choice type of test had slightly higher reliability and seemed slightly easier to score. Hence, evidence from the study supported the use of the multiple choice rather than multiple true-false form, if omissions are not expected.

Data presented by Knowles (1963) demonstrated that questionnaires of the true-false type can be differentially prone to acquiescence response set. This topic is discussed in detail in Chapter X.

It was noted by Tuckman and Lorge (1953) that graduate students experienced frustration because of the either/or choice when circling a yes or no response when asked whether they generally agreed or disagreed with statements about older people. Hence, the authors conducted a study where the same questions were used but where the response was the percentage of older people for whom the question would apply. No significant differences were found between the two methods, causing the authors to conclude that the yes-no method was preferred due to its scoring ease.

Comparison of multiple choice and forced choice or paired comparison items. Appel (1959) administered a 72 item true-false questionnaire and a parallel forced choice questionnaire consisting of 24 triads. The content of the items on each form was identical. Based upon the forecasted validities for the best keys of the two forms for an infinite number of items, Appel concluded that, for longer forms, the forced choice method is likely to result in greater validity, while for shorter forms the true-false method is likely to prove superior.

Osburn, Lubin, Loeffler, and Tye (1954) compared the relative validity of forced choice and single stimulus yes-no self-description items. The contents of the items that were compared were identical. No significant differences were found, but the results seemed to suggest that the choice of format would depend upon the number of items available and their statistical characteristics.

A forced choice form of an interest inventory was compared with a like-indifferent-dislike form by Perry (1955), using the same items for groups of Navy yeomen and college students. Unit weight and multiple weight keys were developed for each inventory to differentiate yeoman from students. The forced choice keys were superior in separating groups in seven of ten comparisons. However, there was little difference in validity shrinkage for the two kinds of items.

Comparison of multiple choice items and card sorts. Van Der Veen, Howard, and Austria (1970) compared response formats of Q-sort, multiple choice, and true-false methods according to test-retest reliability and scoring characteristics. The analyses suggested that all three forms were reliable in test-retest situations. Both the multiple choice and Q-sort methods showed high stability. However, the former showed some variance for social desirability. The true-false method was found psychometrically inferior, showing lower stability and some social desirability variance. The authors concluded that the Q-sort is the format of choice if testing time is available, otherwise the multiple choice format should be used.

Comparison of multiple choice and open-ended items. Two articles compared multiple choice and open-ended items. In the first, Rugg and Cantril (1942) examined the form of the question in public opinion polls by using multiple choice, dichotomous choice, and free response formats. Through five different polls they reached the conclusion that in all cases no one method was best. Multiple choice gave accurate placement, while dichotomous was simply scored. In addition, free response gives respondents the most freedom of expression.

In the second study, Gustav (1964) compared responses to a questionnaire concerning methods of study and preferences for true-false, multiple choice, and essay questions, with actual test scores for 102 undergraduates. True-false items were liked least. A large proportion of the group reported they studied differently for particular types of examinations, and slightly more than half believed they do equally well on all types of tests despite any preferences.

Forced Choice and Paired Comparison Items

This section begins with a review of some issues related to forced choice and paired comparison questionnaire items. Forced choice and paired comparison items are next compared with card sorts and then with check lists. They were compared with ranking items, rating scales, and multiple choice items in earlier sections of this chapter.

Some issues related to forced choice and paired comparison items. In a paper presented by the Personnel Research Section, PRPB, Adjutant-General's Office (1946), it was noted that the utility of rating scales for predictive purposes or administrative action had been limited by the ease with which a rater could determine accurately where he was placing a person on a scale. A technique, the forced choice method, which reduces the rater's ability to control the final result of his ratings, was described. It was noted that the essence of the forced choice technique is to force the rater to choose between descriptive phrases which appear of equal value (have the same preference index) but are different in validity (discrimination index). The major problem is the grouping of alternatives to achieve these ends. The preference index is the mean of the scale indicating the degree to which the phrase applies to the group concerned, while the discrimination index represents the correlation of the descriptive phrase and an overall rating.

As noted by Buel (1963), in the construction of a forced choice scale the preference and discrimination indices are usually derived from responses to items in check list form. The items are then grouped on the basis of similar preference index but dissimilar discrimination index. It is assumed that the preference value of an item does not change when it is transferred from its position in the check list form to a position in the forced choice form. In his study, Buel (1963) found that,while the preference index values of only a few items changed, such shifts generally had the effect of reducing the discrimination index values. Waters and Wherry (1961b) also investigated the stability of preference index values from check list to forced choice administration, and found a high degree of stability. Berkshire and Highland (1953) reported that a favorableness index fits into the forced choice rationale better than does the preference index. Bartlett (1960) made comparisons between the two, and concluded that,if for practical reasons only one index is used for matching, the preference index appears to be the better.

Two studies that were reviewed considered the failure to adequately match the forced choice items. Bartlett (1960) used a scale where the items within each set were not perfectly matched on preference, discrimination favorableness, general factor loading, and magnitude of group factor loading. Multiple correlations indicated that about half of the variance of rating response for both peer and self-ratings could be explained by failure to match on these five indices. Eisenberg (1965) found significantly lower scores when a form was used with items not matched on preference index and different on discrimination index, compared to an identical form developed along classical lines.

Zavala (1965), in his review of the development of the forced choice technique, pointed out that the reliabilities and validities of the technique compare favorably with other methods, and that studies have shown that the method is more resistant than other scales to effects of bias. Earlier, however, Travers (1951) conducted a critical review of the validity and rationale of the forced choice technique and noted that, as used in officer efficiency reports, the evidence did not support claims made for the validity of the procedure. He also concluded that the high validity coefficients secured must be considered to be largely spurious until they are demonstrated to be otherwise. As noted in an earlier section, Scott (1968) similarly concluded that the generalization that self-report forced choice personality inventories are more valid than single stimulus forms of the same tests was not supported by critical consideration of the relevant evidence.

In the study by Berkshire and Highland (1953), six kinds of forced choice formats were compared for rating Air Force technical instructors under experimental conditions and under instructions to give as high a score as possible. The results for the six forms were:

1. Form A: Two statements per block, both favorable or both unfavorable, choose the more descriptive or the least descriptive. Had relatively high reliabilities and validities, was one of the two best liked, but was markedly unsatisfactory in its failure to resist leniency effects. Was also uneconomic in that over half of the blocks failed to discriminate when subjected to item analysis.

2. Form B: Three statements per block, all favorable or unfavorable, choose the most and least descriptive statements in each block. Low in validity, lowest in reliability, least liked by the raters, and uneconomic. Was, however, resistant to skewing under instructions to bias.

3. Form C: Four statements per block, all favorable, choose the two most descriptive statements. Most bias resistant, yielded consistently high validities under various conditions, was one of the two best liked, and had adequate reliability. This method was superior to the other methods tested.

4. Form D: Four statements per block, all favorable, choose the most and least descriptive statements. Comparable to Form C in reliability and validity, but was more susceptible to leniency effects and less well liked.

5. Form E: Four statements per block, two favorable and two unfavorable in appearance, choose the most and least descriptive statements. An inadequate method, easily biased, low validity, and not as well liked as Forms A, C, and F.

6. Form F: Five statements per block, two of which were favorable, one neutral, and two unfavorable in appearance, choose the most and least descriptive. Too easily biased for use. Was moderately well liked, but was exceeded in validity by Forms A, C, and D.

Agreeing with Berkshire and Highland, Zavala (1965) also noted that formats using four favorable items, from which the rater chooses the items most characteristic of the person rated, proved superior to other formats. He said that this superiority appeared in validities, reliabilities, and preferences of raters using the form.

In other studies related to forced choice format, Zuckerman (1952) found that the like-indifferent-dislike arrangement of self-report interest inventories was clearly superior to the two choice form. Waters and Wherry (1961a) reported on the effect of response format on subject resistance to a forced choice self-rating scale. The subjects were found to be more favorable toward a response format allowing them to indicate the degree of applicability of each statement in the forced choice pair, even though they were still forced to choose one statement as relatively more applicable. Waters (1966) also found that reaction to a forced choice scale was more favorable when some method was incorporated whereby the subject was given an opportunity to indicate the degree of applicability of each item to himself.

The effects of partial pairings was investigated in two studies. McCormick and Bachus (1952) conducted a study to determine the extent to which it would be possible, in paired comparison ratings of employees, to use reduced numbers of pairings and still achieve essentially the same rating results as would be obtained from a complete pairing of all individuals within a group. The results showed that ratings obtained from partial pairings resulted in fairly high correlations with ratings based on complete pairings. The correlations were reduced rather systematically with reductions in the number of pairs per individual on which the ratings were based. In a follow-up article, McCormick and Roberts (1952) reported that the reliability of ratings obtained with partial pairings also tended to decrease rather systematically with reductions in the number of pairs per individual on which the ratings were based. However, for groups of 50 individuals, ratings based on as few as 16 pairs per individual appeared to be relatively reliable.

As noted above, Zavala (1965) reported that studies on the forced choice method showed it to be more resistant than other scales to effects of bias. As will be discussed in Chapter X, the forced choice method has been used by a number of investigators in an attempt to control the tendency of individuals to answer self-report inventories in terms of response sets rather than giving "true" responses. For example, Jackson and Minton (1963) concluded that combining items into scales and casting them into a paired comparison context is the method of choice in constructing adjective check lists for personality assessment. This conclusion was based upon the effects of the forced choice format in enhancing content reliability and eliminating the massive response set to check many or few items on the check list. Howe (1960), however, working with anxiety scales, reported that data concerning reliability and skewness did not give an unequivocal impression that the forced choice format reduces the tendency to give socially desirable responses. Feldman and Corah (1960) also reported that social desirability is not minimized by the forced choice format. Braun (1969) pointed out that there is no effective control for social desirability of alternatives presented, nor for fake-proof devices. Lederman (1971) interpreted data as showing that the forced choice format cannot prevent subjects from presenting a more favorable image of themselves if they choose to do so, but that the problem is usually less in the forced choice than in the questionnaire format.

Studies of the results of asking subjects to fake their responses have been conducted by a number of authors, including Izard and Rosenberg (1958) and Eisenberg (1965). Izard and Rosenberg used a forced choice personality test with naval aviation cadets. They found that forced choice scores under instructions for a "set to fake" did not significantly differ from regular scores, suggesting that the test is not easily susceptible to faking. Eisenberg (1965), however, found that instructions to fake did affect results when a forced choice format developed in the classical manner was used.

Comparison of forced choice items and card sorts. In a study by Turgut (1963), a paired comparison format and a modified Q-sort were compared for efficiency in personality measurement. The experiment tested reliabilities per unit of testing time and acceptability to the examinees. The internal consistency coefficients of the paired comparison format averaged .77. The average for the Q-sort was .73 when corrected for the average time spent. Subjects' reaction to the formats were measured by a rating scale, and showed that 57% liked the paired comparison form and 32% liked the Q-sort.

Comparison of forced choice items and check lists. Forced choice or paired comparison items were compared with check lists in two studies. In the U.S. Department of Army study (1952) previously referred to, a controlled check list with 24 items where the 12 most descriptive were to be selected was used in addition to forced choice pairs. The validities of the forced choice pairs, based on rankings by approximately 20 classmates, was .41; of the controlled check list, .44.

Merenda and Clarke (1963) compared two self-rating adjective check lists. The first was the regular free response list, the second a forced choice version where the adjectives were arranged in tetrad sets. Ipsative scoring (discussed in Chapter XI) was used. The results suggested that the forced choice method is likely to be inappropriate for use with adjective check lists in self-concept assessment.

Card Sorts

The advantages of using card sorts for acquiring rating information on any issue has been discussed by a number of authors, including Dubois (1949-50). The most extensive discussion of the use of card sorts (or, more generally, Q-technique and its methodology) probably appears in The Study of Behavior by William Stephenson (1953). Card sorting as a technique for survey interviewing was discussed by Cataldo, Johnson, and Kellstedt (1970), who assessed its reliability, validity, and response bias and the reactions of respondents and interviewers. They concluded that card sorting is a fast and interesting method of obtaining valid and reliable interview data, and one which appears to be capable as well of counteracting at least some of the biasing effects of response set.

Four articles were abstracted that addressed the issue of whether Q-sorting procedures should allow a free or unforced sort where the subject is allowed to place as many cards as desired within the sorting intervals, or require a forced sort where a predetermined number of items have to be placed in each interval cell. Block (1956) compared forced and unforced Q-sorting procedures using 76 items with 11 sorters. The forced sort seemed to offer more stability and slightly more discrimination than unforced sorting. Gaito (1962), considering statistical and non-statistical aspects of Q-sorting, concluded that severe defects appeared present for various analysis tests of significance when forced sorting was involved; moderate distortion when the free sort was used. Hess and Hink (1959) also compared the forced and free Q-sort procedures, and found that the two types of administration gave similar results when the identical Q-sort was used with adolescents. A similar conclusion was reached by Brown (1971). He noted that arguments favoring free over forced Q-sorts had assumed that forcing leads to the loss of important statistical information and interferes with interval properties, rendering Pearson's $\underline{r}$ inappropriate for analysis. He found that Q-sorts with identical item orderings but with varied distributions provided essentially the same correlations and factor structures when coefficients were computed using Spearman's r_s, Kendall's τ, and Pearson's $\underline{r}$. Hence, he concluded that the same results are obtained despite distribution and whether interval or ordinal statistics are used.

In previous sections of this chapter, card sorts were compared with rating scales, multiple choice items, and forced choice items.

Semantic Differential Items

This section reviews some of the pros and cons about the use of the semantic differential, and presents only a few of the many articles on the technique. The first major paper on the semantic differential was by Osgood (1952), in which the development of the technique as a general method of measuring meaning was described. It involved: the use of factor analysis to determine the number and nature of factors entering into semantic description and judgment; and the selection of a set of specific scales corresponding to these factors which can be standardized as a measure of meaning. Using this differential, the meaning of a particular concept to a particular individual can be specified quantitatively. The classical book on the semantic differential was written by Osgood, Suci, and Tannebaum (1957).

Two studies that were reviewed investigated the reliability of the semantic differential. Jenkins, Russell, and Suci (1958) had 360 words rated on 20 scales by 18 groups of 30 subjects. Profiles of mean scale values for each concept were prepared. The reliability of these scale values was found to be .97, and mean scale values correlated .97 with median scale values. Miron (1961) investigated the influence of instructions upon the test-retest reliabilities of the semantic differential, and found the correlations ranged from .996 to .857. The basic measure used in the experiment was the absolute deviation between mean concept scores on each of five factor scores summarizing a given set of scales.

Two reports of the validity of the semantic differential were reviewed, both in the marketing area. Agostini (1962) reported evidence on the validity of the technique as an indicator of brand attitude as measured by purchase behavior. Significantly higher brand average attitude scores were found among users of two brands of a food product than among nonusers, thus illustrating the validity of the semantic differential for this use. Barclay (1964) also found that the semantic differential, in the form used, was a valid indicator of brand attitudes as inferred from purchasing behavior. However, the differential as used was found not to be a very sensitive measure.

Proximity error in administering the semantic differential was studied by Kane (1968). Proximity error occurs when, due to the ordering or polarity of the semantic differential scales, one answer results in another answer to a subsequent question being substantially changed from what it would otherwise be. He investigated effects due to the order of concept presentation, of adjective presentation, and of order of adjectives within a particular scale. He found no significant differences in response traceable to questionnaire format manipulations, showing that proximity error was not a problem with semantic differential questionnaires.

Worthy (1969) noted that semantic differential rating scores are often reported as an extreme response measure which ignores the middle or neutral categories as a response. He reanalyzed data and concluded that those who tended to make extreme responses also tended to make midpoint responses. The implication for scoring was not to make the assumption that a midpoint response is totally lacking in extremeness since it is a demonstrative response. A related concern has been whether or not the semantic differential measures both the intensity and direction of attitude. Mehling (1959) plotted subjects' ratings on an intensity scale against responses to related semantic differential scales, and concluded that as used in the study the semantic differential did measure both the direction and intensity of attitude. Bentler (1969) also found that semantic space is approximately bipolar, while Carter, Ruggels, and Chaffee (1968) found that subjects can more accurately denote their descriptions to objects when one end of the scale is left for them to describe.

Semantic differential scales were compared with check lists by Block (1958) and Hughes (1967). Block found a correlation of .94, after correctio for attenuation, between semantic differential descriptions and adjective check list descriptions of the ideal self and the liked-sex parent. He concluded that the semantic differential may be a rather complicated way of developing a measure that is more readily and reliably secured by other means. Hughes (1967) reported that 28% of the semantic differential scales he used detected attitude change, while only 11% of the check list scales did. Both showed the same test-retest reliability, however (.58). Preference for the semantic differential increased from 11% to 34% as the respondents became familiar with it, while preference for the check list declined from 57% to 40%.

Comparisons of the semantic differential with other types of rating scales appear in an earlier section of this chapter.

Other Types of Items

The types of items to be considered in this section are projective items, open-ended items, check lists, rearrangement items, and matching items. Comparisons of check list items and rating scale, forced choice, and semantic differential items were discussed in previous sections of this chapter.

Projective items. The use of projective items was not a high priority topic for this report, but three reviewed documents discussed them. In the study of attitudes of troops in the tropics authored by Hart, Faust, Rowland, and Lucier (1964), complementary objective and projective techniques were compared and contrasted for their efficacy in assessing attitudes towards items of QM issue and situations relating to tropical military service. They found that projective and unstructured data collection techniques provided attitudinal data not captured by the more structured techniques. They also found that responses to objective items correlated significantly with sentence completion items on the same topic. Thematic stimuli provided in a projective pictures-written response technique were, however, inadequate for eliciting the appropriate topic related attitudinal responses. A color response technique did not indicate any relationships with other techniques. Nevertheless, the authors recommended that a combination of highly structured, semistructured, and unstructured techniques be employed in a complex measurement setting, as is typical in the case of attitude measurement.

In the marketing area, Haire (1950) found that projective methods may aid in determining respondent's motivations toward a stimulus in linking attitudes and behavior. Steele (1964) investigated the validity of projective questions, and concluded that the projective technique is a useful device where inhibitions may be raised in an interview.

Open-ended items. A comparison of open- and close-ended questionnaire items was presented in Chapter II, while the use of open-ended items to determine questionnaire content was discussed in the first section of this chapter.

On other relevant topics, Roslow, Wulfeck, and Corby (1940) noted that results from free response questions may be misleading when the memory of the respondent and/or familiarity with possible responses operates to any appreciable extent. Payne (1965) cites a meaningful role for open-ended questions in preliminary phases of research in areas such as the development of categorical, checkbox questions, to eliminate the need for asking reason-why questions of every respondent, or to provide quotes which may add interest to a report. Frisbie and Sudman (1968) found a direct relation between the amount of speech in open-ended questions and positive and negative feelings. People with high positive or negative feelings talked more than those with low positive or negative feelings. In both cases, those classified as having high feelings had one more sentence than those classified as low.

Comparison of open-ended items and check lists. Two studies compared the use of open-ended questions with check lists. Scates and Yoemans (1950b) studied the effect of question form on the course requests that were received from adults employed in scientific and engineering fields. They found that questionnaires involving depth essay questions were returned by a smaller proportion of persons, but the requests which they contained were believed to be more firmly based. A course check list elicited a larger number of course requests per employee who returned it than did questions which asked the employee to think of the courses he may desire.

The check list and open response methods of survey research were compared by Belson and Duncan (1962) with respect to yesterday's reading of newspapers and magazines, and with respect to yesterday's TV viewing. Results indicated that offering items in the form of a check list produced an appreciably higher rate of claim that publications were looked at. However, the check list was found to depress the enumeration of items placed under its "other" category. The open response system produced only 73% of the volume of endorsements produced by the check list, but it gave 1.72 times as many compared to the "other" category.

Other topics regarding the use of check lists. Roslow, Wulfeck, and Corby (1940) found that the proportions obtained by alternatives in check list questions tended to be influenced by the number and completeness of the alternatives presented. And Lindzey and Guest (1951) found that omissions of popular items from check lists produced substantial changes in response distribution. They also found that few respondents used the "other-write in" category.

McCormick (1960) studied the effect that the number of questions asked about each task had on the consistency and amount of information provided by Air Force personnel when completing task inventories. No systematic differences were found in the number of tasks reported by incumbents who were asked to report one, two, three, or four types of information about such tasks. The requirement to report more types of information generally provided more reliable information.

Rearrangement items. Sims (1934) examined the use of rearrangement tests of ability as an alternative to objective tests. He found that as the length or the rearrangement set increased from five to 15 items, the reliability also increased. At some point before 30 items, however, the length of the test seemed merely to reflect the student's intelligence. He concluded that this type of test compares favorably with other types of objective tests in reliability, time for taking, and scoring time, when the desire is to measure the ability to relate items to some designated basis.

Articles more pertinent to the use of rearrangement items in questionnaires were not located during the literature search which preceded this review.

Matching items. The literature review uncovered only one article on the topic of item matching. Follman, Urbanke, and Burley (1971) compared three item matching objective test formats using 60 undergraduate students who were asked to match definitions to 20 appropriate verbs typically used in essay type questions. The results indicated that better scores were obtained by ordering the items randomly but dividing them into small groups of three to six items.

Conclusions Regarding the Pros and Cons of Various Types of Questionnaire Items

Ranking items. Based upon five studies, it would appear that ranking and rating techniques are generally comparable. There is some evidence, however, that conclusions based upon a single judge differ from those based upon multiple judges. More research appears to be needed in this area, especially studies designed so that the items to be ranked or rated are as comparable as possible.

Contradictory evidence was obtained regarding the comparison of ranking and paired comparisons. The bulk of the evidence, however, seems to support the notion that the two techniques produce comparable results. In two studies a linear relationship was found between the results obtained from the two. Several investigators noted that, if the results are comparable, ranking is to be preferred to paired comparisons since it takes less time.

Rating scale items. A majority of the studies reviewed found that results obtained from the use of rating scales were comparable to those when forced choice or paired comparison items were employed. This is surprising in terms of the known properties and limitations of the ipsative scores produced by forced choice devices, as discussed in Chapter XI. However, Scott (1968) noted that generalizations about forced choice and single stimulus tests are open to question since most studies did not use identical items in the two forms compared. If results from the two are comparable, then rating scales might be preferred since they are more efficient and take less testing time.

Results comparing rating scales and card sorts are inconclusive. Only two studies were found in this area. Results are similarly inconclusive in the comparison of rating scales and semantic differential items and rating scales and check lists, as few studies were located.

Multiple choice items. Comparisons between multiple choice items and true-false items (which are a special type of multiple choice item) are discussed in Chapter VI. The three studies that reported comparisons of multiple choice and forced choice items did not come to the same conclusion, although there was some tendency to feel that a choice of format would depend upon the number of items available and their statistical characteristics. In the one study located that compared multiple choice items and card sorts (Van Der Veen, Howard, & Austria, 1970), it was concluded that the true-false method was psychometrically inferior, and the Q-sort should be used in preference to the multiple choice format if adequate testing time is available. There is also little on which to base a conclusion regarding multiple choice and open-ended items. Each might have its place, depending upon the purposes and objectives of any given study.

Forced choice and paired comparison items. Some issues related to forced choice and paired comparison items were reviewed. Based upon two studies, problems seem to arise when the alternatives within a forced choice item are not adequately matched. Although one investigator (Zavala, 1965) pointed out that the forced choice technique compared favorably with other methods in terms of reliability and validity, two others (Travers, 1951; Scott, 1968) felt that some of the claims made were not supported by the evidence. A more critical and detailed review of the studies conducted in this area is probably in order.

Two investigations (Berkshire & Highland, 1953; Zavala, 1965) lead to the conclusion that the best format for forced choice items (at least when used for personnel ratings) is four statements per block, all favorable, where the two most descriptive statements are to be chosen. Zuckerman (1952), however, found three statements preferable to two. In terms of what was noted regarding the need to adequately match items, more research on this issue would probably be worthwhile.

There still seems to be some question as to the extent that forced choice items can be used to reduce undesirable response sets, at least in terms of the articles included in this review. Since this was not a topic of great stress during the literature review because many of the articles are in the personality area, a more intensive literature review on the use of forced choice items to control response set might be in order.

Only one study was found comparing forced choice items and card sorts, while only two were located comparing forced choice items and check lists. Conclusions about them would not appear warranted.

Card sorts. A majority of the articles about card sorts addressed the question of whether or not forced or unforced sorts should be used. The conclusion appears to be that when the same items are used, it does not make much difference which system is employed.

Semantic differential items. There are a number of investigators that advocate the use of the semantic differential, and its reliability and validity seem to have been established in the articles reviewed. Block (1958), however, questioned whether the semantic differential may be a rather complicated way of developing a measure that is more readily and reliably secured by other means. Since the technique was initially developed as a general method of measuring meaning, a more extensive literature review regarding its use in questionnaires might be in order.

Other types of items. Projective items, open-ended items, check lists, rearrangement items, and matching items were also discussed above. Since there were few studies about these types of items, conclusions do not appear warranted.

Chapter IV

COMPARISON OF SCALING TECHNIQUES

Once a selection has been made of the kinds of items that are to be used in a questionnaire, it may be necessary to determine scale values for the items. Chapter IV is addressed, therefore, to a comparison of scaling techniques. The literature review made in conjunction with preparing this report, however, did not stress articles on scaling techniques, although some were uncovered. Since the topic was not stressed and many other articles could be located, a discussion similar to those included in other chapters does not appear warranted. Instead, the articles for which abstracts were available from the literature search are listed below with a short annotation regarding their content. Comparisons of psychological scaling techniques with other types of questionnaire items are discussed in Chapter III in those sections pertaining to rating scales.

Ballin and Farnsworth (1941) developed a graphic rating method that had scale values which agreed closely with scale values obtained using the Seashore-Hevner method and the method of equal appearing intervals.

Banta (1961) used the methods of Likert, Thurstone, and a newly developed method of Unfold Partial Rank Order to measure attitudes towards each of three referents differing in ambiguity. The scores obtained from all three methods correlated equally well at each level of referent ambiguity.

Barclay and Weaver (1962) found that the construction of a Thurstone scale took 43.2% more time than the construction of a Likert scale with the same number of items, and that the Likert scale was more reliable.

Bartlett, Heermann and Rettig (1960) compared the magnetic board rating technique to the paired compairson, ranking, Likert, graphic rating and equal appearing intervals methods. It was concluded that all six scaling techniques were equally accurate measures of scale value.

Clark and Kriedt (1948) applied Guttman's scaling techniques to the Rundquist-Sletto attitude scale. The scale did not meet Guttman's criterion for adequate scale undimensionality despite the fact that the internal consistency of the scale was high. Thus the authors concluded that Guttman's method of scale analysis may have serious limitations in the area of general attitude measurement.

Coombs (1950) developed the ordered metric scale which is based on the order of magnitude of the interval between objects.

Edwards (1946b) concluded that Likert scales tended to have higher reliability than Thurstone scales and were easier to construct.

Edwards (1948) refined Guttman's technique for determining cutting points by assuming perfect reproducibility and making predictions of item responses on this assumption.

Edwards (1951) discussed the use of the method of successive intervals, which is a psychological scaling procedure in which stimuli are classified into successive intervals according to the degree of some defined attribute which they are judged to possess.

Edwards (1956) concluded that using the method of paired comparisons in conjunction with a set of opinion statements with known scale values had promise for the construction of attitude scales with a relatively high degree of reproducibility and satisfactory reliability.

Edwards and Kenny (1946) established the fact that it is possible to construct scales by the Likert and Thurstone methods which will yield comparable scores.

Edwards and Kilpatrick (1948b) described the Scale-Discrimination method which makes use of Thurstone scaling procedures, retains Likert's process for evaluating the discriminatory power of the individual items, and meets the requirements of Guttman's Scale Analysis.

Eysenck and Crown (1949) handled the results of a study by: determining reliabilities under various systems of scoring (Thurstone, Likert and Scale Product); factorial analysis; Guttman Scalogram analysis; plotting of scale positions of items against number of endorsements, percentage reproducibility, and factor saturations; and determining neutral point by different methods.

Farnsworth (1945a) found that scale weights obtained for items using a technique modified from Allport approximated weights obtained by Thurstone with sorting. The modified technique was a method where extreme items were put at the opposite ends of a series of equilength lines representing the individual items and where the subjects, in a group situation, checked relative item value.

Farnsworth (1945b) as the result of a study where judges of statements were asked their understanding of the distance between degrees of the scale, questioned the use of equal appearing interval scales.

Federico (1971a) studied Likert and Guttman-type questionnaire forms. He found that Air Force students demonstrated significantly more favorable attitudes toward analogous content areas on the Guttman-structured items than on the Likert-structured items. Evidently, item formatting did affect the degree of the evaluative assertions ascribed to the attitude universe.

Ferguson (1939b) suggested the following requirements for attitudinal scales: scale results correspond to underlying physical order; scale values selected not affected by other items in scale; attitudes of judges of responses do not affect scale values; specific in content; validity; reliability; and scale on a linear continuum.

Ford (1950) illustrated a rapid method for determining whether a set of six (or fewer) attitude questions form a scale.

Gardner (1950) suggested a technique for obtaining an interval scale not dependent on an assumed normal distribution or the selection of one given population. The units of this scale are called K-units.

Guilford (1928) presented a method for getting scale values which are assumed to be on an "objective" continuum. These scale values are in terms of sigma units from an assumed mean of all the stimulus values.

Guttman (1947a) described the Cornell technique, which is primarily the combining of data to produce cutting points that minimize error of reproducibility.

Gulliksen & Messick (1969) included in their book discussions on: the method of successive intervals; quantitative judgment scales; similarity of stimuli; metric properties of behavioral data; the method of successive categories; ratio scales; partition scales; confusion scales; and multidimensional unfolding.

Hughes (1967) compared the Thurstone scale, a modification of the semantic differential, and the check list scales for their ability to detect changes in attitudes, their test-retest reliability, and their acceptance among respondents.

Jahn (1951) extended scale analysis along three lines; one, to include alternative methods for reduction of a set of attributes to a single quantitatively defined variable; two, to include methods for the reduction of a set of attributes to a single qualitatively defined variable or qualitative types; and three, the development of statistical-experimental tests to decide whether the theorems of scale analysis are to be accepted for application to a given empirically defined set of attributes.

Kelley, Hovland, Schwartz and Abelson (1955) found that data analysis using the method of equal appearing intervals did not discriminate judges with extreme views.

Komorita (1963) demonstrated a neutral region could be determined for Likert scores but because of the quasi-scale characteristic of the instrument no neutral point could be clearly delineated. Weighting content scores by intensity as in the Likert method, instead of using simple zero-one weights, had negligible effects on total score. However, if the number of items is small, there seemed to be some advantage in the Likert method.

Kriedt and Clark (1949) concluded that the Cornell Technique of Scale Analysis (Guttman) can prove to be very useful in problems of psychological measurement providing discretion is exercised in the selection of suitable problems and the handling methods.

Kundu (1962) modified the scale points on a Likert scale. This modification was a factor dividing rating method whereby the neutral point is eliminated on theoretical grounds and the remaining scale points are not fixed in advance by the test author, but are assigned weights by the respondent in accordance with his prevailing response bias.

Likert (1932) presented the background and theory for his measurement approaches.

Prothro (1955) found data that supported Thurstone's assumptions that the sorting of items into an attitude scale is independent of the attitude of judges.

Rozeboom & Jones (1956) stated that the degree to which scale values computed by the method of successive intervals diverge from theoretically "true" values is seen to be due to three types of error: error due to inequalities in variances of the distribution from which the scale values are computed; error due to nonnormality of the distribution; and sampling error.

Saffir (1937) made a comparison between scales constructed by the method of paired comparison, rank order, and the method of successive intervals. He found mutually linear scales, and concluded that all the methods he employed produced equally valid scales. Since the three different methods of gathering data (method of paired comparisons, order of merit method, and method of successive intervals) and the two different psychophysical techniques for scaling raw data (the law of comparative judgment and the method of successive intervals) produced comparable scales, any one can be used with considerable confidence.

Schaie (1963) hypothesized that the concurrent validity of questionnaires could be increased by the use of item weights obtained by expert scaling, instead of by using conventional unit weights. The results showed only low magnitude increments in validity.

Seashore and Hevner (1933) modified Thurstone's method of equal appearing intervals by having judges rate items on a nine point scale which was printed on the left hand margin of each item, instead of sorting items printed on separate slips into nine piles.

Siegel and Schultz (1962)demonstrated that a job related technical skills check list could be scaled by both Thurstone and Guttman techniques.

Siegel, Schultz, and Benson (1960) hypothesized that skills are scalable in the same manner (Guttman and Thurstone equal appearing interval scales) as attitudes and sensory phenomena. Although their results supported the hypothesis, discrepant data raised some question as to the generality of the hypothesis.

Siegel and Siegel (1962) found that medians graphically derived and medians from sorted judgments scaled by the method of equal appearing intervals correlated .97.

Sjöberg (1965) compared conventional scaling techniques with his correlational scaling method using paired comparisons data. Nonlinear relations between scales were found.

Stangenberg (1966) presented definitions of various scales (the nominal, ordinal, interval, ratio and logarithm) and discussed them in terms of measurement theory.

Stouffer, Guttman, Suchman, Lazarsfeld, Star, and Clausen (1949) defined in their book the components of a scale and discussed the limitations of the use of scales. Their work was a result of studies carried out with military subjects during World War II.

Taylor and Parker (1964) found that the graphic rating scales proved as reliable as Guttman scales, and an examination of the interscale correlations showed that similar conclusions would be drawn from either technique.

Thurstone (1959) presented the background and theory for his measurement approaches.

Torgerson (1958) presented in his book definitions and explanations of scaling methods. The book includes extensive numerical examples.

Witroyl (1954) experimentally compared Thurstone's Case III and Case V and Guildord's shortcut approaches to scaling paired comparison data. The intercorrelation between the scale values obtained by the three methods were approximately unity.

York (1966) found that Thurstone's scale values are stable over 35 years.

Zinnes (1969) reviewed the literature on scaling. The theme of the review was that scaling theory should be a theory of choice.

Chapter V

EFFECTS OF VARIATION IN PRESENTATION OF QUESTIONNAIRE ITEMS

Once a decision has been made regarding the type or types of items that are to be used in a questionnaire based on the pros and cons discussed in Chapter III, attention must be given to the actual development of the items. In this chapter consideration is given to articles in the literature that investigated the effects of variations in the presentation of questionnaire items. Sections are included on the: mode of items; wording of items; clarity of items; difficulty of items; length of question stem; order of question stems; and order of response alternatives.

Mode of Items

A series of research studies were uncovered concerning verbal versus pictorial presentation of items/stimuli for subject's responses. The studies covered a variety of topical areas and types of subjects. Table V-1 summarizes the literature review conducted in this area.

Four studies found no significant differences in subjects' responses to verbal and pictorial formats (Blake, 1969; Greenberg, 1959; Jensen, 1930; and Rohila, Shanhdhar & Sharma, 1966). Only one study, relating to a consumer preferences examination,showed statistically significant differences attributed to mode of item presentation (Weitz, 1950). It should be noted that this study did not establish superiority of one format over the other, but merely noted differences on brand ratings. Another study on the influence of communications (Luchins and Luchins, 1955b) provided an important screening procedure for the use of pictures in questionnaire items. This study suggested that conformity with false communications and failure to respond were higher for ambiguous than clear-cut pictures. Obviously, if pictures are to be used they should be pretested for clarity of their presentation of the concept or object to be evaluated.

The overall evaluation of this area of the literature is that pictures can be effectively employed in questionnaires. This may facilitate obtaining survey responses from subjects with limited verbal comprehension who might have difficulty responding to questions employing lengthy definitions of concepts or objects.

Wording of Items

The wording of question stems and response alternatives is a critical consideration in obtaining valid, reliable, and objective survey data. For example, Payne (1951) cited the following illustration. The three questions following were administered to three separately matched samples of respondents (Payne, 1951).

TABLE V-1

Summary of Studies on Mode of Items

Study	Topic	Verbal vs. Nonverbal Treatment	No.subjects/type	Results
Greenberg (1959)	Psych. profiles of new products	4 picture stereotypes of female users vs. verbal description	250/housewives	No difference between 2 sets of associations for individual brands
Blake (1969)	Occupational interest inventories	Picture vs. verbal descriptions	192/males	Small tendency for picture forms to yield higher reliabilities (not strong nor consistent over items)
Jensen (1930)	True-false examinations	Visual, oral, & visual-oral format	25/college students	All three reliabilities high and essentially equal, mean test scores similar
Luchins and Luchins (1955)	Influence of communications	Ambiguous vs. nonambiguous pictures	144/college students	Conformity with false communication and failures to respond were higher for the ambiguous than for the clearcut pictures
Rohila, Shankhdhar & Sharma (1966)	Preference measures	Picture (nonlanguage preference record) vs. verbal (Kuderlike triads)	301/boys	Changes in item form did not produce significant differences in group means although a number of significant differences appear on individual items

(Table continued on next page)

TABLE V-1 (cont.)

Summary of Studies on Modes of Items

Study	Topic	Verbal vs. Nonverbal Treatment	No.subjects/type	Results
Weitz (1950)	Consumer ratings of cooking ranges	Rank attributes (design, layout, color, etc.) on a verbal questionnaire vs. same attributes via pictures	100/females	Significant differences on overall brand ratings existed between verbal and picture groups

1. Do you think anything should be done to make it easier for people to pay doctor or hospital bills? (82% replied "yes")

2. Do you think anything could be done to make it easier for people to pay doctor or hospital bills? (77% replied "yes")

3. Do you think anything might be done to make it easier for people to pay doctor or hospital bills? (63% replied "yes")

These questions differed only in the use of the words should, could and might, terms that are often used as synonyms even though they have different connotations. The 19% difference at the extremes is enough to alter almost any survey's conclusions. This example illustrates a key feature of much of the evidence on question wording -- it is extensively topic bound. Most of the studies dealing with framing questions were so broad in scope that no single source of bias was given concentrated attention (Belkin and Lieberman, 1967; Hubbard, 1950). It is difficult to generalize from the literature to a specific survey situation.

The literature review conducted for this section uncovered several articles and books purporting to offer "principles of question wording" (e.g., Payne, 1951; Roslow & Blankenship, 1939; Blankenship, 1942). Most of the material presented, however, is based on experience rather than empirical research, and tends to be more prescriptive than positive, more indicative than imperative.

The stress in the discussion of the literature presented below is on topics which have been discussed in some detail: positive versus negative wording of items; objective versus subjective wording of items; and definite versus indefinite article wording. A section on miscellaneous studies on questionnaire wording has also been provided to list areas which have been examined as "one-shot" efforts.

Positive versus negative wording of items. One topic in question wording which has received considerable attention is statement polarity, positively versus negatively phrased question stems. Table V-2 summarizes the literature on this topic. It should be noted that all the studies except Adams (1956) concern question stem wording. Only three studies (Adams, 1956; Githens, undated; and Waters, 1966) were unable to find an effect on study results produced by positive versus negative wording. Eleven studies reported significant effects on a variety of measures, such as reliability, validity, and suggestibility (Blankenship, 1940a; Burtt & Gaskill, 1932; Campbell, Siegman & Rees, 1967; Cloud & Vaughn, 1970; Edrich, 1965; Falthzik & Jolson, 1974; Hubbard, 1950; Muscio, 1916; Rugg, 1941; Rundquist, 1940; and Wembridge & Means, 1918). In general these studies produced evidence that alternative positive/negative (or neutral wordings) can produce demonstrable effects on survey results -- a conclusion not arguing for either form of phrasing but mere recognition that differences in results existed.

TABLE V-2

Summary of Research on Positive Versus Negative Wording of Items

Study	Topic area	Pos./Neg. treatment	No.Subjects/Type	Results
Adams (1956)	Attitudes toward city commercial facilities	Asked if things were better, the same or worse vs. better, same or better elsewhere	341/residents of Winston-Salem	No significance on 8 of 9 items (respondents didn't inhibit using strongly unfavorable response alternatives)
Blankenship (1940b)	Opinion of government policies	Positive (obj. & subj.) and negative (obj. & subj.) questions (among other treatments)	3,000/residents	Negative objective questions 2nd most valid; positive objective questions most consistent (small difference)
Burtt and Gaskill (1932)	Questions about newsreel	Negative vs. positive form of question	1,090/college students	Negative caused greater suggestiveness (% following lead questions) when categorical (yes, no, don't know) answers demanded
Cambell, Siegman & Rees (1967)	Personality tests	Items stated positively, negatively or neutrally	unknown/unknown	Direction-of-wording effect present in all scales examined
Cloud and Vaughn (1970)	Conservatism scale	Balanced positively keyed and negatively keyed items	1,831/college students	Balancing reduced yea-saying tendencies

Summary of Research on Positive Versus Negative Wording of Items

Study	Topic Area	Pos./Neg treatment	No.Subjects/Type	Results
Edrich (1965)	Attitude toward Negroes	One version high % of unfavorable items vs. balanced Likert items	Unknown/unknown	Scale values significantly lower with greater unfavorable version
Falthzik and Jolson (1974)	Attitude toward unit pricing	Varied positively/ negatively stated Likert items	486/supermarket shoppers	Statement polarity significant on 7 of 12 items (higher agreement with positively posed than disagreement with negatively posed, especially in personalized items); Acquiescence inversely related to respondents' education
Githens (undated)	Attitude toward Navy career	3 versions: pos. aspects neg. aspects pos./neg. aspects	Unknown/officers (NROTC)	No effect
Hubbard (1950)	Unknown	Neg./neutral(among others)	unknown/unknown	Introducing a negative into question form increases suggestibility and decreases reliability (personalized version increases suggestibility vs. nonpersonalized)

(Table continued on next page)

TABLE V-2 (cont.)

Summary of Research on Positive Versus Negative Wording of Items

Study	Topic area	Pos./Neg. treatment	No.Subjects/Type	Results
Muscio (1916)	Observations of behavior	Positive versus negative items with subjective/objective treatments	56/College students	Interaction effect -- the most reliable form was a subjective directed question without negatives or definite articles
Payne (1951)	Doctor and hospital bills	Should vs. could vs. might	Unknown/unknown	82% said yes with should, 77% said yes with could, 63% said yes with might,
Rugg (1941)	Freedom of speech	Allow vs. forbid (speeches against democracy)	Unknown/unknown	62% said no when asked allow, 54% said no when asked forbid
Rundquist	Personality tests	Positive versus negative items	Unknown/eighth grade students	Items of like-type correlated higher than unlike (ideational content); Negative items tend to intercorrelate to a greater extent particularly for girls
Waters (1966)	Unknown	Positive, negative or neutral items	108/college students	No effect
Wembridge and Means (1918)	Voting measures	Simple affirmative, simple negative (prohibit or not), double negative, i.e. "minors shall not be forced not to smoke"	Unknown/unknown	Test subjects took greater time and were more confused in answering double negative than simple affirmative questions

Several additional similarities in results among the studies were present. Both Burtt and Gaskill (1932) and Hubbard (1950) reported that introducing a negative into the question form increased respondent suggestibility. That is, there was a tendency for the direction of the question stem to be chosen in the response alternatives. A potentially important interviewing variable was pointed out by the comparable findings of Falthzik and Jolson (1974) and Hubbard (1950). These two studies illustrated a tendency for statement polarity to be more significant when personalized (what a person says about himself) than nonpersonalized (what he says about others or external events). Furthermore, their results indicated that when a personalized question was changed to a nonpersonalized version, suggestibility was decreased. Two other studies cast doubt on the use of negatives in question stems. Muscio (1916), in assessing the reliability of subjects reporting events they had just observed, reported that the most reliable question form was a subjective directed question without negatives. In a study of appropriate question stems for voting measures, Wembridge & Means (1918) reported that respondents took greater time and were more confused with negative and especially double negative (i.e. minors should not be forced not to smoke) stems than simple affirmative versions. These findings are contradicted by Payne (1951) who reported from several studies that, when people have strong convictions, the wording of the statement should not greatly change the stand they take. Rundquist (1940) also suggested that negative items in a series of personality measures tended to have greater internal consistency than positively phrased items.

In conclusion, loading by statement polarity choice may be unavoidable but can cause differences in research results. It can even be desirable when evaluating policies or objects. But when a particular phrasing is employed to present a distorted view of opinion or the view in which the researcher thinks is "right," it becomes an evasion of truth, or the direct opposite of research (Payne, 1951).

Objective versus subjective wording of items. Eight studies were uncovered relating to the effects of stating question stems in an objective or subjective direction. A study published by Muscio (1916) is illustrative of the research in this area. Fifty-six subjects were exposed to a sequence of pictures and then asked if they saw certain objects in them. The study's dependent variable was suggestibility, or the degree to which subjects said "yes" to these objects whether they were present or not. Muscio concluded that changing from the subjective ("Did you see a hat in the picture?") to the objective ("Was there a hat in the picture?") reduced suggestibility. Table V-3 summarizes additional evidence in this area.

Muscio's evidence regarding objective-subjective direction and suggestibility has been supported by empirical studies conducted by Blankenship (1940a), Dohrenwend (1965), and Hubbard (1950). The only conflicting evidence uncovered in the area was presented by Burtt and Gaskill (1932), who reported that the objective form showed greater suggestibility.

TABLE V-3

Summary of Research on Objective Versus Subjective Wording of Items

Study	Topic Area	Objective/subjective treatments	No.Subjects/Type	Results
Blankenship (1940a)	Legalized betting	5 forms of question varying degree of subj. to obj.	3,000/N.J. residents	Predictive accuracy and reliability highest for most objective version (Is it desirable to permit legalized betting in N.J.?)
Blankenship (1940b)	Government policy	Objective vs. subjective wording (Do you....) on yes, no, don't know responses	3,000/town residents	Objective most valid; Objective least suggestive; Objective most indecision (don't knows appropriate here)
Burtt and Gaskill (1932)	Evaluation of a motion picture	Objective vs. subjective forms	1,090/college students	Objective form showed greater suggestiveness and higher number of "don't knows"
Dohrenwend (1965)	Campus interest item	Objective vs. subjective interview questions	32/college students	Objective version had higher response specificity, correctness; subjective version led to greater self-revelation, and specificity of identification of other persons
Fiske (1969)	Personality inventories	First person vs. third person wording	300/airmen	Significant differences between wording "What would others say about you" had higher scale values than self-description

(Table continued on next page)

TABLE V-3 (cont.)

Summary of Research on Objective Versus Subjective Wording of Items

Study	Topic Area	Objective/subjective treatments	No. Subjects/Type	Results
Hubbard (1950)	Attitudes towards a brand of kitchen utensils	Subjective (you) vs. objective direction (it) of wording	Unknown/unknown	Changing a subjective direction into an objective direction form decreased suggestibility and reliability
North and Schmid (1960)	Job satisfaction	Personal vs. impersonal	199/airmen	Personal form had highest internal consistency and test re-test reliability
Muscio (1916)	Objects seen in pictures	Subjective vs. objective	56/unknown	Changing from the subjective to objective phrasing reduced suggestibility

Concerning response specificity or the avoidance of the "Don't know" category, three studies with conflicting results were found. Dohrenwend (1965) noted that the objective version had higher response specificity, while Burtt and Gaskill (1932) and Blankenship (1940a) concluded that "don't knows" increased with the objective direction.

The reliability of objectively or subjectively phrased questions has also been investigated. The only study presenting evidence of higher reliability for objective question versions was presented by Blankenship (1940a). Hubbard (1950) and North and Schmid (1960) presented evidence that subjectively phrased items were more reliable.

The only research studies reporting validity evidence have been conducted by Blankenship (1940a, 1940c). In both studies the objectively stated version of a question had higher predictive accuracy.

It seems that follow-up research is warranted in this area. The limited research evidence points up more contradictions than similarities in findings. The only area where a tentative conclusion favoring objective over subjective phrasing can be made is in the area of suggestibility.

Definite versus indefinite article wording. Two studies were found which reached similar conclusions regarding the use of definite or indefinite articles in question stems. Indefinite article ("a" or "an") items are exemplified by the following type of question -- "Did you see a demonstration of the new night vision device?" A definite article ("the") item would be worded -- "Did you see the demonstration of the new night vision device?" Studies by Muscio (1916) and Hubbard (1950) both concluded that changing from "a" to "the" wordings reduced the level of suggestibility. The use of indefinite article questions, however, led to increased reliability of answers when factual or objective information was sought (Hubbard, 1950). No conclusions in this area can be drawn because of the limited evidence available.

Miscellaneous studies on question wording. The previous three areas, positive versus negative, objective versus subjective, and definite versus indefinite article wordings, have been researched in a somewhat systematic fashion. This section, however, is designed to present selected highly relevant studies about question wording which have not been replicated by other social scientists.

Several isolated studies have dealt with the effect of building into the question some reference to prominent people. For instance, Cantril (1940a) compared responses to the following two questions: "Do you approve of President Roosevelt's sending Sumner Welles to visit European capitals?" and "Do you approve Sumner Welles's visit to European capitals?" When Roosevelt's name was used, more people had opinions and more people disapproved (25% versus 31%). In other studies of the suspected "big name effect" the results have varied, the big name sometimes making a difference and sometimes not (Belson, undated b).

Research has also been conducted on the consequences of employing stereotypes, emotion charged, or culturally biased words. Significant changes in the frequencies of positive, negative, or don't know responses may result (Roslow, Wulfeck & Corby, 1940).

The potpourri of scattered studies on question wording are illustrated by the following:

1. Hubbard (1950) reported that the incomplete disjunction form, e.g., "Was the demonstration interesting, dull or just so-so?" possessed relatively high suggestibility and relatively low reliability.

2. North and Schmid (1960) examined all possible combinations of personal-impersonal and qualified-unqualified forms of questions, and concluded that personal-qualified versions may be the best according to internal and test-retest reliability and independence criteria.

3. Steele (1964) compared projective, direct, and indirect questions. Information from the projective item (interpretations of a picture drawing) explained more variance in the dependent variable (milk consumption) than direct and indirect questions.

4. Waters (1966) presented evidence that a subject's reaction to a forced choice scale was more favorable when some method was incorporated whereby the subject was given the opportunity to indicate the degree of applicability of each item to himself. "Most descriptive of you" and "least descriptive of you" were not as effective as "the degree to which this applies to you" on a five point scale.

5. Thumin (1962) experimentally examined buffer items in question sequencing. Buffer items were defined as neutral items intended to establish rapport which were placed before "delicate" items. Study results indicated that the buffer items increased respondent's admissions of insomnia.

6. In a study involving return of a job satisfaction questionnaire by 47% of over 1,000 life insurance agents, the effectiveness of direct and indirect questioning techniques was examined (Weitz and Nuckols, 1953). Results indicated that the direct and indirect items intercorrelated significantly, but that the direct items in general had greater validity (were better predictors of job survival).

7. Richardson (1960) tested the widespread assumption that interviewers should not use leading questions. Tape recorded interviews of seven experienced and 30 untrained interviewers were compared. Questions were classified into leading and nonleading questions. A leading question was operationally defined as one which includes, either explicitly or implicitly, the answer which the interviewer

expected to receive. Study results showed that: contrary to expectation, experienced interviewers used leading questions in 33% of all their questions; and leading questions elicited no more responses containing distorted information than did non-leading questions.

Conclusions. The literature discussed in the area of question wording illustrates certain gaps and thinnesses. Many important issues have been raised by these studies, but there has been little systematic pursuit of the issues to a conclusion. Seldom have replication studies been uncovered or attempts made to carry them out in different settings. Many of the studies have been carried out with small samples or only with college student subjects. Finally, no research studies were uncovered which examined the wording of response alternatives.

Clarity of Items

Almost every check list of prescriptives for writing question items includes instructions such as: "make sure your questions are clear to the respondent;" "avoid ambiguous, vague, and imprecise questions;" and "questions must be concrete and specific." Unfortunately, there has been little systematic research on measures of clarity or the effects of unclear questions on subjects' responses. In fact, most of the available literature is based upon authors' experiences, intuition, and "common" sense (e.g., Jenkins, 1941; Roslow & Blankenship, 1939). This issue has received some investigation in the area of interviewer ambiguity in phrasing questions (Hanson & Marks, 1958), a topic not covered here.

This section will present the limited evidence available on how to improve item clarity and the effects of item clarity on subjects' responses.

Studies on improving item clarity. Several interesting studies have been reported which suggested diverse tactics of improving item clarity. Gray (1955) suggested that in framing questions which depend on respondents' memory or recall capabilities, the time period a question covers must be carefully defined and redefined. The when should be specifically provided. Litwak (1956) suggested that ad hoc rules on question wording, such as cautions against loaded, vague, double-barreled questions, can be investigated by latent structure analysis. Evidence was presented that bias in questions may lie in too many (ambiguity), too few (clarity), or inappropriate (clarity) dimensions. Thus, alternative question wording and additional descriptions may aid a subject's interpretation. Toops (1937) reported that subjects preferred a format where key words or portions of the question stem were capitalized. The inference was that an idea of what is required to respond to the question and overall clarity can be obtained by a glance at the capitalized material. No results were reported concerning whether underlining key words might accomplish the same goal.

Several studies were found which are attempts to isolate the amount of clarity in questions. Speak (1967) conducted a study whereby subjects who had responded to questions in a personal interview were reinterviewed the following day by another "in-depth" session to ascertain what the respondent had "really" meant and how he interpreted the questions. It was found that not one question was perceived by every subject as intended, nor did one subject perceive all the questions as intended. It appears then that follow-up interviews might be purposeful in screening paper and pencil question items. For example, Nuckols (1953) submitted poll questions to respondents and then, after completion, asked them to interpret the meaning of the questions. At least 17% of these interpretations were judged to be wholly or partially wrong.

Another clarity screening method has been offered by Norman (1963b). He conducted a study of test item content in personality measurement. The results indicated that there existed marked differences in the validities obtainable from different classes of test stimuli, those with the highest degree of judged content relevance producing the most satisfactory results. To the degree that relevance enhances the clarity of questions, this would also seem to be an appropriate pretesting procedure.

A technique called the "random probe" was used to check what closed questions actually meant to respondents in a survey in Pakistan (Schuman, 1966). Interviewers were instructed to select randomly 10 items for further probing. Respondents' understanding was then ranked on a five point scale. Results indicated that with this particular instrument a significant minority of the respondents had real difficulty with the questions.

Miklich (1966) studied response sets in relation to ambiguously worded statements. Forty-two subjects were given statements with four types of treatment: ambiguous, unambiguous, important, and unimportant. They were asked whether they agreed or disagreed. The analysis indicated that ambiguous items did result in more agreement-disagreement response set. That is, if the ambiguous item was important (not defined in the study writeup) the tendency was to agree with it, while if unimportant, the tendency was to disagree.

A large scale study (Belson, undated c) of respondent understanding of over 2,000 items used in market and social survey questions included a content analysis of reinterview data with 265 subjects regarding their understanding of the original questions. Findings related to item clarity were: if a broad term or concept was used in a question, there was a strong tendency for respondents to interpret it less broadly; and respondents who failed to hear some part of a question tended to reconstruct the question from what they had heard.

Effect of item clarity on subjects' responses. Few studies have been conducted in this area, perhaps because question clarity is itself such a vague, general concept. One important study was offered by Armstrong and

Overton (1971). Two versions of a questionnaire about intentions to use a new transportation service were tested. One version using a brief description and one using a comprehensive description were successively administered. No significant differences were found on estimates of level of demand at various prices, or on the identity of likely user groups. Thus, in some cases, additional verbal material in questions or topic descriptions may not alter subjects' responses.

Difficulty of Items

One of the first "laws" of questionnaire development advanced by almost every general source on how to write sound questionnaires is the statement "keep it simple." Logic dictates that words used in surveys should not have multiple meanings, nor should they be beyond the level of vocabulary of the typical respondent. Unfortunately, this advice is often poorly operationalized.

This section discusses measures of item difficulty, and miscellaneous studies of survey instruments. The abstracted literature on item difficulty is summarized in Table V-4.

Measures of item difficulty. A series of studies have taken standardized tests or published public opinion poll questions and subjected them to a form of content analysis against reading or vocabulary difficulty indices. Payne (1950a) found that "tightly worded" questions on an opinion poll had Flesch scores at 7th or 8th grade level whereas "loose" questions with large variance in reverse worded items scored at the high school level or above.

Similarly, Nuckols (1953) reported that nine published poll questions had remarkable problems in wording difficulty. In an independent retest, 17% of his subjects had interpretations of individual questions which were judged partially or wholly wrong. Flesch scores ranged from 5.8 to 17.2 in reading grade. Another study (Terris, 1949), again a reexamination of poll questions, compared Flesch and Dale-Chall readability scores to Census Bureau Reports on formal school levels of the U.S. population. Study results indicated that:

1. 91.6% of all the questions were above the comprehension level of 12.4% of the population.

2. 73.4% of all the questions were above the comprehension level of 23.2% of the population.

3. 9.8% of all the questions were above the comprehension level of 72.6% of the population.

Difficulty of items has also been assessed with The Teacher's Word Book of 30,000 Words (Thorndike and Lorge, 1944). Users of this source state that it is best to err on the side of simplicity if doubt exists.

TABLE V-4

Summary of the Literature on Item Difficulty

Study	Topic Area	Types of Instrument/Items	No.Subjects/Type	Results
Faeber (1951)	Timed arithmetic test	Open answer, right-wrong, multiple choice, multiple choice with separate answer sheet	230/6th and 7th grade students	Forms showed increasing difficulty in order given
Hanley (1965)	Personality test	3 measures of item difficulty used: controversiality, response latency, and confidence in accuracy of answer	Unknown/unknown	Acquiescence occured with difficult rather than easy inventory material. Response latency and subject's confidence superior to controversiality as difficulty measure
Myers (1962)	Educational tests	Homogeneous vs. heterogeneous item difficulty	Unknown/unknown	No difference in validity coefficients, homogeneous difficulty tests more reliable
Nuckols (1953)	Public opinion	Nine published poll questions	48/unknown	17% of subjects' interpretations judged wholly or partially wrong. Flesch scores ranged from 5.8 to 17.2 in reading grade.
Payne (1950)	National poll on petroleum tax	Unknown	6400/national cross-section sample	"Tight" questions (differences in answers to reverse-worded items less than 2%) have characteristics of brevity and simplicity. Flesch scores at 7th or 8th grade for "tight" vs. high school level or more for "loose"

(Table continued on next page)

TABLE V-4 (cont.)

Summary of the Literature on Item Difficulty

Study	Topic Area	Types of Instrument/Items	No.Subjects/Type	Results
Roeber (1948)	Interest inventories	Seven of most popular interest inventories	Not applicable	10% to 20% of vocabulary used in each test was above 9th grade reading level as measured by Thorndike and Lorge word list
Stricker (1963)	Attitude and personality measures	Varied readability in items	93/students	Easy-to-read items elicited least acquiescence while hard-to-read or moderate items most acquiescence. This relationship was found for attitude items but opposite for personality items
Terris (1949)	Public opinion polls	Used Flesch and Dale-Chall formulas for readability. Compared results to Census Bureau reports on formal school levels.		91.6% of all the questions above comprehension level of 12.4% of population; 73.4% of all the questions above comprehension level of 23.2% of population; 9.8% of all the questions above comprehension level of 72.6% of population.

There are many examples of misunderstandings of what seem to be everyday words. One study (Roeber, 1948) found that 10% to 20% of the vocabulary used in seven of the most popular interest inventories was above the 9th grade reading level as measured by the Thorndike and Lorge word list.

Unfortunately, no studies were found concerning the comparative results of reading level measures using scores on the Flesch, Dale-Chall, Thorndike and Lorge, or other readability scales. Also, no information was uncovered regarding the "fog level" reading difficulty scoring system used by the U. S. Air Force, or Fry's Readability Graph (Fry, 1968). However, a detailed literature search in these areas was outside the scope of this review.

Hanley (1965) suggested two other measures of item difficulty with reference to personality testing: response latency and subjective confidence in accuracy of answer. Either of these measures might also be employed in pilot or pretest studies of survey instruments.

Miscellaneous studies of survey instruments. More attention has evidently been devoted to measures of item difficulty than to the effects of item difficulty on questionnaire responses. Hanley (1965) and Stricker (1963) offer the exceptions. Both have examined the impact of item difficulty on acquiescence response bias, but with conflicting conclusions. Stricker determined that acquiescence was more prevalent with moderate or hard-to-read attitude items, but found the opposite relationship for personality items. Using the response latency and subjective confidence difficulty measures, Hanley (1965) concluded that acquiescence occurred with difficult, rather than easy, personality inventory material. Additional attention obviously needs to be focused on item difficulty in terms of response tendencies such as acquiescence, "don't know" and "no" responses.

One study (Faerber,1951) has addressed the important matter of comparative difficulty of different response alternative formats. In a timed arithmetic test, open answer, right-wrong, multiple choice, and multiple choice with separate answer sheet formats were experimentally manipulated. Results showed increasing difficulty in the order listed here.

Finally, Myers (1962) compared homogeneous to heterogeneous item difficulty educational tests. No difference in validity coefficients were found, but tests homogeneous in difficulty were shown to be more reliable.

Length of Question Stem

Only three studies were found in the literature search effort which dealt with length of question stems. It should be noted, however, that the topic of instrument length is discussed in Chapter VIII.

The first study located (Brinkmeier, 1930) was only marginally relevant to questionnaire construction because it concerned true-false tests for high school student examinations. A statistical analysis of 6,671 question stems submitted by high school teachers in a national contest revealed that stems under twenty words in length were as often true as false. As length increased beyond twenty words, however, the probability that the answer was true increased.

Marquis, Cannell, & Laurent, (1972) examined the impact of question length and respondent education on self-reports of health information. The data were later compared to physicians' reports of the same information for each respondent. Results indicated that longer (interview) questions increased the accuracy of reports from those who had finished high school, and had the opposite effect on those who had not. Another study (Laurent, 1972), perhaps reporting on the same data base as the previous citation, offered evidence drawn from four experiments conducted with samples ranging from 24 to 200 interviews for the U. S. Public Health Service. Questions were altered in length by adding redundant, inconsequential information in various treatments. It was found that the longer questions elicited more information than the shorter questions. Also, after checking with physicians' reports, it was found the longer questions received more accurate answers.

This is apparently an underresearched area. The few isolated studies just reviewed concern only objective tests and interview schedule development. The conclusion that longer question stems (controlling for age) produce a greater amount and more accurate information cannot be generalized based upon the limited evidence. More research is warranted in this area.

Order of Question Stems

Several different sources of error must be considered regarding the general issue of question stem effects in questionnaire methodology. Order bias has several meanings in survey research concerning question stems. For example, if the question were asked, "Which kind of weapon do you prefer, the M14 or the M16," one might conjecture that a reversing of the order of alternatives within the question stem might be a source of response error. Literature in this area is discussed in the section of this chapter on the wording of items. Order bias in this section refers to the order of questions within a series of items designed to explore the same subject matter, or related subject matter areas. A related issue concerns the position effect problem -- the order of different groups of questions, when the groups deal with essentially unrelated subject matter areas.

Table V-5 presents a summarization of literature dealing with the order bias and position effect problems.

TABLE V-5

Summary of Studies Relating to the Order of Question Stems

Study	Topic	Type of Order Effect Examined	Order Treatment	No.Subjects/Type	Results
Baehr (1953)	Job attitudes	Order bias	Randomized vs. categorized items	452/department store employees	No difference in order
Bardburn and Mason (1964)	Mental health	Position effect	Sections rotated (work experience, marriage & family, social participation, etc.)	2,789/unknown	No difference in order
Blumberg DeSoto, & Kuethe (1966)	Rating well-known names on traits	Order bias	1 name at a time vs. 1 trait at a time vs. free choice order	100/unknown	No difference in order
Brenner (1964)	Difficulty of test items	Order Bias	Varied item difficulty order	Unknown/unknown	Differently ordered forms (difficulty had no effect on test difficulty or reliability)
Cohen (1965)	Unknown	Order bias and position effect	Varied order of questions within a series of the same or related subject matter; varied groups of unrelated questions	Unknown/unknown	Results were dependent on order-differences of as much as 15% (response distribution on items). No position effect.

(Table continued on next page)

TABLE V-5 (cont.)

Summary of Studies Relating to the Order of Question Stems

Study	Topic	Type of Order Effect Examined	Order Treatment	No.Subjects/Type	Results
Ferber (1966)	Consumers Union questionnaire	Order bias	Analyzed existing order vs. item non-response	14,600/subscribers	No relationship between frequency of item non-response and position of question on the questionnaire
Gross (1964)	Interest in buying a new pen	Order bias	Five sequences tested: interest alone; product advantages-interest; product disadvantages-interest; advantages-disadvantages-interest; disadvantages-advantages-interest	Unknown/unknown	Buying interest lower when disadvantages considered first
Hofstee (1966)	Trait descriptive adjectives	"Secular trends -- systematic changes in response to test items resulting from experience with similar items	Investigated trend for items to be responded to more desirably as a list continues and trend for items to be responded to more or less desirably depending on desirability of the preceding item	576 Dutch/ military men	No evidence for first trend. Second trend confirmed -- responses more desirable if preceding term more desirable
Landon (1971)	Ratings of professors	Position effect	Rated 'ideal'professor 1st then instructors vs. reversed order	441/graduate students	Instructors were rated significantly lower when ideal rated first

(Table continued on next page)

TABLE V-5 (cont.)

Summary of Studies Relating to the Order of Question Stems

Study	Topic	Type of Order Effect Examined	Order Treatment	No.Subjects/Type	Results
Lyman (1949)	School Attitude Inventory	Order bias and position effect	Items scrambled vs. items blocked according to scale	400/high school seniors	No differences between scrambled and blocked
Metzner and Mann (1953)	Unknown	Position effect	Sequence form: related topical questions grouped vs. random order	884/employees	Results inconclusive: within group response interrelationships increased 7 times in sequence form, 3 times in random, 10 times no significant correlation
O'Dell (1962)	Comparison of personal interviews and mail surveys	Sequence bias	Compared empirical data from a number of studies	Unknown/unknown	The ordering of questions was shown to effect the accuracy of responses in mail panels (not being able to see the whole set of questions in advance - minimizes sequence bias in personal interviews)
Survey Research Centre (1970)	Brand preferences	Position effect	Two sequences tested of semantic differential ratings; (1)each brand rated on attribute 1, then on attribute 2, etc.; (2) first brand on attributes 1, ...12, the 2nd brand on attributes 1...12, etc.	1,400/London, England voters	Significant differences from presentation orders (rating sequences) were: higher average endorsements under (2); don't know responses greater in (2); (2) gave a mean score close to the estimated accurate mean (from an earlier study)

(Table continued on next page)

TABLE V-5 (cont.)

Summary of Studies Relating to the Order of Questions Stems

Study	Topic	Type of Order Effect Examined	Order Treatment	No.Subjects/Type	Results
Survey Research Centre (1972)	Brand preferences	Order bias	Rotated order of brands rated on semantic differential scales	1,068 London, England voters	For individual brands, the order of presentation could influence: distribution of ratings; average number of ratings on a single scale; and the number of "don't know" ratings
Symonds (1936)	Adolescent problems and interests	Order bias	Reversed order of items to be ranked to two groups	975/junior high students	No effect

Order bias. One of the most typical caveats discussed in the general literature of how to construct questionnaires is the statement, "vary (randomly assign) the order of questions on an instrument to avoid one question contaminating another." Especially prominent are the discussions of cases where the immediately preceding question or group of questions places the respondent in a different "mental set" or frame of reference. For example, asking respondents a general question about their feelings about automobile exhaust pollution might influence responses to a question like: "Do you prefer leaded or nonleaded gasoline?" Although this effect may be prominent in specific, applied settings, little evidence was found in the literature supporting a general, order bias phenomenon in survey research. Five studies (Baehr, 1953; Blumberg, DeSoto & Kuethe, 1966; Brenner, 1964; Ferber, 1966; and Lyman, 1949) were unable to document order biases in investigations in divergent topical areas. Five studies (Cohen, 1965; Gross, 1964; Hofstee, 1966; O'Dell, 1962; and Survey Research Centre, 1972) found that the presentation order of question stems significantly affected response distributions to given items, nonresponse to items, and preferences for specific stimuli. Thus, the findings in this area are inconclusive -- no support exists for the presence of a general order effect in questionnaire responses. Although the literature that was reviewed on this topic was sparse, it appears that order effects are a function of the specific instrument and subjects employed in the investigation. It is interesting to note that order bias or question sequence may be a subtle issue in specific cases. More experiments testing the effect of changing the sequence of questions have been uncovered that show no effect than show significant differences.

Position effect. Practical advice on how to avoid position bias problems abounds in the questionnaire development literature. Suggestions to phrase questions in a logical sequence, build rapport first, ask for the basic information sought next, and personal questions last, are illustrative of the guidelines offered the questionnaire designer. From the literature review, it appears that the extent of a general position bias is unknown. This is an area that is poorly documented. Four studies (Bradburn & Mason, 1964; Cohen, 1965; Lyman, 1949; and Metzner & Mann, 1953) were unable to find any effect of changing the sequence in which major sections of questionnaires were presented. Conflicting evidence was offered by experimental results presented by Landon (1971) and the Survey Research Centre (1970). Again, it must be concluded that systematic research in this area is lacking. As in the previous case, however, position bias may be operative in specific research situations, but the weight of the evidence supports a negligible influence of position bias on survey findings.

Conclusions. The results in the areas of order bias and position effect cannot be regarded as definitive. In light of the unknown in these areas, individual questions and question sections can probably be placed into whatever appears to be the best psychological or most logical order.

Order of Response Alternatives

One of the principles of questionnaire development advanced by psychologists is that the responses to a particular proposition will be influenced by the position of the alternative in the question. In the literature of questionnaire methodology, it is also known as the "time error" and can occur in questionnaire applications as well as with laboratory methods. Mathews (1929), in one of the earlier works to recognize this response pattern, noted in reviewing the results of an experimental study that, although overall differences in sequencing existed, the first of two alternatives in a question where the order was varied received more endorsements than the second position. This study also suggested that the fourth (of five) response alternatives was chosen somewhat more frequently. Mathews' work has received only token empirical support with respect to other reviewed literature. Belson (1965) and Winthrop (1958) offered evidence that reversal of verbal or numeric rating scale response alternatives are coupled with a significant shift in endorsements toward the first presented end items or anchors. Belson (1965) reported that a reversal from positive to negative scale orders resulted in a greater proportion of choices of negative (or unfavorable) end categories. Winthrop's evidence suggests, similarly, that reversal of numerical preference alternatives in natural numbers order (e.g., 1, 2, ..., 5) results in lower scale reliability.

Two additional studies documenting an order effect in response alternatives were found. Becker (1964) reported that subjects' choices of their five favorite types of radio and T.V. programming were influenced by the ordinal position of the choices in a checklist. This study suggests that, as an item is listed close to the end of a checklist, the probability of its selection is reduced. Madden and Bourdon (1963) found that reversing the order of levels of job factors that were presented to airmen for evaluation of various jobs resulted in significant differences in job ratings.

The studies discussed above must be regarded as the exceptions rather than the rule in this research area. Seven experimental studies (Blumberg, DeSoto & Kuethe, 1966; Campbell & Mohr, 1950; Clark, 1956; Dyer, Klein, & Yudowitch, 1975; Feldman, 1969; Kane, 1971; and Symonds, 1936) reported little or no order effects with response alternatives. The first study, for example, experimentally manipulated the "good" end of a graphic rating scale in left, right, top, or bottom positions with minimal resultant effect on ratings. The analyses conducted by Dyer, Klein, and Yudowitch (1975) concerned a VOLAR study administered at Fort Hood, Texas, with over 500 military subjects. Reversal of response alternatives was accomplished by presenting one-half of the subjects with alternatives listed from most positive to most negative; e.g., "The training I have received at Fort Hood has been: very challenging, challenging, borderline, unchallenging, very unchallinging." The remaining subjects received response alternatives listed from most negative to most positive. This treatment, used on both attitude and satisfaction scales in the VOLAR questionnaire, did not produce significant differences on either individual items or categories of items.

Several problems uncovered by the literature reviewed for this section preclude arriving at any valid generalization concerning order effects in response alternatives. First, most of the available published studies have been conducted with relatively small samples of college student subjects. Second, the number of studies conducted in this area is limited. Third, no systematic research has been published with respect to the order of response alternatives in specific types of rating or scaling devices, such as graphic or verbal ratings, semantic differential, or Likert scales. Fourth, important moderating variables such as subjects' characteristics, topical area, scale length (number of response alternatives), and instrument length have neither been controlled nor built in as experimental treatments.

The reviewed studies on the order of response laternatives are summarized in Table V-6. Because of the inconclusive nature of the findings and their contradictions, care probably should be taken to alternate the order of response alternatives when it appears appropriate to do so. In this vein, it should be noted that several authors (Ross, 1934; Mosier & Price, 1945) have developed tables to standardize the order of presentation of words. Although these tables were designed to provide systematic variation of paired comparison and multiple choice items, they may be applicable to verbal rating scales as well. Ross (1934) states that his method will aid in wording "regular" repetition patterns, providing optimum spacing between identical words, and balancing out fatigue effects.

TABLE V-6

Summary of Studies on Order of Response Alternatives

Study	Topic	Method of Order Treatment	No.Subjects/Type	Results
Becker (1954)	Preferences in radio & T.V. programming (news, sports, music, etc.)	Varied ordinal position of 16 categories in check list	193/residents of two states	As an item is listed close to the end of a check list, the probability is lower that someone will select it as one of his five favorites
Belson (1965)	Opinion on television via verbal rating scales	Two groups: (1) favorable or positive items presented first (2) reversed order	332/London, England residents	Reversal of presentation order of response alts. ((1) to (2)), shifted responses toward negative or unfavorable end of scale; first presented end-items (anchors) received greater endorsement
Blumberg, DeSoto & Kuethe (1966)	Rated well-known names on traits	The "good" end of a graphic rating scale was at the left, right, top, or bottom	100/unknown	Location of positive end of a scale with numeric or graphic rating has minimal effect on ratings
Campbell and Mohr, (1950)	Preferences for types of radio programming	Varied ordinal position of the categories in check list	1,280/college students	No significant differences in preferences (average rank of top five picked due to order
Clark (1956)	College administration tests	Examined position effects of response alternatives in five choice tests	unknown/college applicants	Positional preferences are weak

(Table continued on next page)

TABLE V-6 (cont.)

Summary of Studies on Order of Response Alternatives

Study	Topic	Method of Order Treatment	No.Subjects/Type	Results
Dyer, Klein, & Yudowitch (1975)	Volunteer Army Program Evalution	Two groups: (1) favorable or positive alternatives presented first; (2) reversed order	500/officer and enlisted military personnel	Reversal of order did not aff either attitude or satisfacti ratings
Feldman (1969)	Ratings of adjectives and adjective parts (semantic differential)	Examined position effects of adjectives and combinations	75/college students	The order of presentation doe not affect the ratings of adjective combinations
Kane (1971)	Various types of order effects in semantic differential scales	Varied scale polarity e.g. reversed order of extreme ends of sem. diff. scales	150/college students	No significant differences du to reversing polar extremes
Madden and Bourdon (1963)	Job evaluation ratings	Reversed order of levels of 9 job requirement factors that subjects used to rate 15 jobs	60/basic airmen	Significant differences due t order existed
Mathews (1929)	Attitudes of students toward study habits	Reverse order of presentation of "like greatly, like, indifferent, dislike and dislike greatly" rating to 2 groups	184/junior high students 130/college students	No significant differences, b the first and fourth response alternative chosen somewhat m frequently

(Table continued on next page)

TABLE V-6 (cont.)

Summary of Studies on Order of Response Alternatives

Study	Topic	Method of Order Treatment	No.Subject/Type	Results
Symonds (1936)	Ranking of problems and interests of adolescents	Items listed in reversed order to two groups	975/students, junior high thru college	No significant difference due to order of presentation
Winthrop (1958)	Preference ratings for fruits	3 groups tested: normal order scale (1, 2, 3...etc.); reversal order scales; normal followed by reversal	60/unknown	Reliabilities of the three scales are in the following order: normal, reversed, mixed

Chapter VI

NUMBER OF RESPONSE ALTERNATIVES AND RESPONSE ANCHORING

The effects of variation in the presentation of questionnaire items, including the order of response alternatives, was discussed in Chapter III. This chapter considers two related topics: the number of response alternatives to employ; and response anchoring.

Issues Regarding Number of Response Alternatives to Employ

One of the basic issues in the use of any given rating instrument or attitude scaling device is the determination of the optimum number of response alternatives or categories. Researcher's habit, or tradition rather than solid empirical support, often has led to the recurrent use of five point Likert scales, seven point semantic differential scales, and so on. The reason for concern with the number of response alternatives stems from the belief that a "coarse" scale with too few response alternatives may result in a loss of information concerning subjects' discrimination powers, or reduced cooperation in rating reflecting a dislike for "forcing" a judgment. An extremely "fine" scale, with too many response alternatives, may go beyond the subjects' powers of discrimination, be excessively time consuming, or difficult to score.

The literature search in the area of number of response alternatives was very productive. Over 30 studies were found which were directly related to this issue. Table VI-1 summarizes the literature. The final three columns, headed Reliability, Validity, and Other Findings, illustrate that multiple criteria have been used in investigating the issue of the optimum number of response alternatives to employ. The major criteria used have been reliability, validity, factors influencing subjects' motivation and ability to respond, and scoring and data analysis considerations. Each of these criteria will be discussed below.

Reliability. Numerous studies (Bendig, 1953; Bendig, 1954a; Komarita & Graham, 1965; Jacoby & Matell, 1971; Masters, 1973; Matell & Jacoby, 1971; Saunders & Ward, 1964) in the psychometric literature have shown that increasing th number of response alternatives does not necessarily increase reliability. These empirical efforts have employed a wide variety of response alternative combinations as experimental treatments, for example, two through 19 alternatives inclusively, and were conducted using several types of rating scales in the context of numerous topical areas. It should be noted that all the studies above except the two Jacoby and Matell efforts were concerned with internal consistency measures, that is, equivalent forms or split-half reliability. Jacoby and Matell (1971) examined both internal and temporal (test-retest) consistency and concluded that both measures were independent of the number of response alternatives.

TABLE VI-1

Summary of Studies Relating
To Number of Response Alternatives

Study	No. of Response Alternatives Varied	Type of Scale/Topic	No. of Subjects/Type	Reliability Internal/Test-Retest	Validity Concurred/Predictive	Other Findings
Baehr (1953)	3, 5	Unknown/job attitudes	454/department store employees			No significant differences between 3 and 5 alternate scales
Bendig (1953a)	3, 5, 7, 9 or 11	Rating scales (anchored and unanchored)/ knowledge of foreign nations (self-rating)	225/college	Reliability equal for 3, 5, 7, and 9 point scales. Reliability decreased for 11 alternatives		
Bendig (1953b)	3, 5, 7, 9 or 11	Same as above Bendig study	Same as above			Increase in amount of information transmitted (more descriptive data) as number of alternatives increased
Bendig (1954a)	2, 3, 5, 7 or 9	Rating scales/ food preferences	236/college	Test reliability (summed ratings for each S) constant over all alternatives. Rater reliability (summed ratings for each food) constant from 5-9, but slightly lower at 2 and slightly higher at 3		
Bendig (1954b)	2, 3, 5, 7 or 9	Same as above	Same as above	Same as above		
Bendig (1954c)	2, 3, 4 or 5	Rating scales/ food preferences	278/unknown	Rating reliability highest with 4 alternatives, lowest with 2 alternatives. Rater bias did not appear to be related to scale length		
Bevan & Avant (1968)	2, 4, 8, 16, 32, 64	Unknown/size judgments	Unknown			1) Response latency (time) increased as a direct function of number of alternatives. 2) Response uncertainty increased up to 32 alternatives but not after

(Table continued on next page)

TABLE VI-1 (cont.)

Summary of Studies Relating
To Number of Response Alternatives

Study	No. of Response Alternatives Varied	Type of Scale/Topic	No. of Subjects/Type	Reliability Internal/Test-Retest	Validity Concurred/Predictive	Other Findings
Dunette, Aylward & Uphoff (1956)	5 (undecided midpoint) vs. 6 (undecided midpoint = I don't know)	Agree-Disagree/ union activities	214 and 216/ union members			On those questions checked "undecided" by 5 alternatives group, almost all had checked "I don't know" by 6 alternatives group.
Cardner (1958)	2-8	Multiple choice /unknown	192/infantry men			The relative frequency of the response to the most frequently presented stimulus object was a function of the number of choices
Ghiselli (1939)	2 vs. 4	Yes-no vs. very sincere-very insincere/sincerity of advertising	102 and 98/ college students			Significantly higher don't knows (no responses) to yes-no scale (willingness to respond). Yes-no subjects rated advertising as less sincere than 4 alternatives group (no statistical tests)
Geldsamt (1972)	6, 12 or 18	Graphic rating scales with verbal anchors/ content free stimuli (geometric shapes, etc.)	120/unknown			Extreme response style was unaffected by number of scale alternatives. Dichotomous scoring as intensity scoring did not affect results.
Green & Rao (1970)	2-18	Not applicable	Simulated data			If reproducibility of the original data configuration is a criterion, 6 to 7 alternatives are optimal
Guest (1962)	2 vs. 4	Agree-disagree vs. SA, A, D, SD/Attitudes towards progressive education	95 and 86/adult residents			Small majority of 2 alternatives group against progressive education vs. majority of 4 alternatives for progressive education (significant differences found on only 1 of 10 items

(Table continued on next page)

TABLE VI-1 (cont.)

Summary of Studies Relating
To Number of Response Alternatives

Study	No. of Response Alternatives Varied	Type of Scale/Topic	No. of Subjects/Type	Reliability Internal/Test-Retest	Validity Concurred/Predictive	Other Findings
Hake & Garner (1951)	1-5, 1-10, 1-20, 1-50	Use of pointer dial to represent response values	16/college males	No difference in mean square error for 10, 20 or 50 pointer positions		
Hughes (1969)	7 alternatives on bi-polar adjective scale vs. 8 (added don't knows alternatives)	Semantic differential scale unknown	84/college			Forced choice scale had piling at midpoint (confounded difference and awareness). When midpoint responses removed no significant differences existed on items.
Jacoby & Matell (1971)	2-19	Rating scales/ values	360 college	Internal consistency of test-retest independent of number of alternatives	Validity independent of number of alternatives	Can score dichotomously or trichotomously according to direction of response
Komerita (1965)	2 vs. 6	Semantic differential scale/ sociability	Unknown/unknown	Increasing the number of alternatives for heterogeneous scale improved reliability but no effect for homogeneous scale (note no main effect for change in number of alternatives)		
Lehmann & Holbert (1972)	3, 5, 7 or 9	Not applicable	Simulated data			The larger the number of alternatives the smaller the rounding error (in summing). If interested in averages (across people or scales), 2 or 3 alternatives permissable.
Lindzey & Guest (1951)	4 vs. 7	Check list/ unknown	42/college			Deleting popular items from check list produced substantial changes in response distribution
Masters (1973)	2-7	Likert/Educational Progressivism	420/graduate students	Reliability of Progressivism scale increased from 2 to 3 alternatives. Reliability of Traditionalism scale independent of number of alternatives.		

(Table continued on next page)

TABLE VI-1 (cont.)

Summary of Studies Relating
To Number of Response Alternatives

Study	No. of Response Alternatives Varied	Type of Scale/Topic	No. of Subjects/Type	Reliability Internal/Test-Retest	Validity Concurred/Predictive	Other Findings
Matell (1970)	2-19	Likert/unknown	Unknown/college			Time to administer was equal. As number of alternatives increased number of uncertain responses decreased. Scales of 9-13 alternatives were most preferred. Collapsing alternatives into 2 or 3 categories resulted in no lack of precision.
Matell & Jacoby (1972)	2-19	Likert/unknown	360/unknown			Proportion of scale used independent of number of alternatives. Mean testing time increased and use of "uncertain" alternatives decreased as number of alternatives increased
McCormick (1952a)	?	Paired Comparisons/employer rating	100/industrial foremen	Reduction in number of pairs (partial comparisons) results highly correlated, with complete pairings. Reliability of partial pairings decreased with reductions in the number of pairs per individual on which ratings based. For groups of 50 as few as 16 pairs per individual "fairly" reliable.		
McCormick (1952b)	8-24	Paired Comparisons/employer rating	See above	Same as above		
Neidt (1951)	2 (positive-negative) vs. 5 (1-5 rating)	Rating scales/attitude toward education	201/college	Reliability of 5 alternatives somewhat higher than 2 alternatives	No difference in validity	
Saunders & Ward (1964)	2 vs. multiple choice	Bipolar scales/personality	282/college	No difference in reliability		Efficiency - no difference in the proportion of positive responses

(Table continued on next page)

TABLE VI-1 (cont.)

Summary of Studies Relating
To Number of Response Alternatives

Study	No. of Response Alternatives Varied	Type of Scale/Topic	No. of Subjects/Type	Reliability Internal/Test-Retest	Validity Concurred/Predictive	Other Findings
Strahan (1971)	2 T.F. vs. several alternatives	Rating scales/ personality	200/college			Ss prefer > number of response alternatives to T-F
Symonds (1924)	? vs. 7	Rating scales/ personality	Results based on formula grinding (Kelly's formula for correcting a corr.coef. for coarseness of grouping)	Fewer than 7 alternatives resulted in loss of reliability, greater than 7 don't improve reliability		
Tsudzuki (1953)	2 (yes-no) vs. 4 (agree-disagree)	Ratings scales/ unknown	Unknown/unknown			Using 4 alternatives format reduced % of nonresponse (preference)
Tuckman & Lorge (1953)	2 (yes or N.R.) vs. 3 (3rd alt. asked for % of people to whom statement applies)	Rating scales/ attitudes about older people	147/graduate students and 51/students			No significant differences in % of yes answers for the 2 methods
Tversky (1964)	Unknown	Multiple choice/ testing	NA-form without grinding			3 Alternatives multiple choice have greatest discriminability, power and information
Zuckerman (1952)	2 (forced choice) vs. 3 (Like, Indifferent, Dislike)	Rating scales/ interest inventories	418 and 94/ college			Like-Indifferent-Dislike superior to forced choice

Several examples of studies exist in the literature which indicate that a nonlinear relationship exists between the range of response alternatives and the magnitude of the coefficient of reliability. For example, in separate investigations Bendig, (1953 , 1954a, 1954c) found: reliability was equal for three, five, seven and nine point scales, but lower for 11 alternatives; rater reliability (summed ratings for each object rated) was constant from five to nine alternatives but slightly higher at three and slightly lower at two categories; and rater reliability was highest with four alternatives and lowest with two alternatives. Neidt and Merrill (1951) reported that the reliability of a five point rating scale was slightly higher than a two alternative, positive-negative, format. In a personality assessment study, Symonds (1924) contended that fewer than seven scale alternatives resulted in a loss of reliability, but employing greater than seven did not improve reliability.

The above studies seem to suggest that there is an optimal number of response alternatives to employ for any given investigation situation, including the topic area, characteristics of the subjects, etc. For example, in other studies results were dependent on the type of rating instrument used. Komorita and Graham (1965) found that increasing the number of response alternatives improved reliability for heterogeneous scales with dissimilar item content, but had no effect on homogeneous scales. Masters (1973) demonstrated that the reliability of a traditionalism of education scale was independent of the number of response alternatives, but in a progressivism scale reliability increased from two to three alternatives.

Validity. Jacoby and Matell (1971) pointed out that most of the psychometric literature dealing with the number of alternative issues emphasized reliability as the major, and in some cases, only criterion in the choice of the number of scale points. They felt, however, that the ultimate criterion is the effect a change in the number of scale points has on the validity of the scale. Table VI-1 illustrates that only two original studies addressed the validity question. Neidt & Merrill (1951) in an attitude toward education investigation reported no difference in concurrent validity between two and five alternative rating scales. This study reported mean course marks and scale scores, holding constant ACE scores and hours studied per week.

The only study examining both concurrent validity, with attitudes and behavior measured at one point in time, and predictive validity (correlation of observed behavior with that which was predicted from attitude measures) was Jacoby and Matell (1971). In relation to both measures, the authors concluded that no consistent relationship existed between either measure and number of response alternatives employed.

Although the evidence is consistent, the lack of numerous studies using divergent types of subjects, instruments and topics, makes it difficult to reach a conclusion regarding the effect of the number of response alternatives on validity.

Factors influencing subjects' motivation to respond and efficiency of response. This section addresses a series of related matters: subject's preferences for and ability to use scales with a varied number of response alternatives. Direct and indirect measures of subject preferences and motivation have been examined. Direct measures, preferences ratings, were used in studies conducted by Matell (1970) and Strahan (1971). However, both studies used college students as subjects. Experimental results were consistent: college students reported a preference for using finer scales. In Matell's investigation, scales of nine to thirteen alternatives were preferred, and Strahan reported significantly higher preferences for using "several" alternatives over a true-false format.

Indirect or proxy measures of subject preferences and motivation include response time and number of "uncertain" and "no responses." Matell (1970) presented evidence that no difference in total time for administration was shown for two through nineteen alternatives. Bevan and Avant (1968) and Matell and Jacoby (1972), however, reported that testing time increased as a direct function of the number of alternatives, thus supporting the more intuitively plausible findings. Concerning the relationship between the use of "uncertain" response categories or "no responses" and scale length, the literature supported the conclusion that increasing the number of response alternatives decreased uncertain and non-responses to scale items (Matell, 1970; Hughes, 1969; Matell & Jacoby, 1972; Ghiselli, 1939; Dunette, Alyward & Uphoff, 1956; Tsudzuki, 1953; Zuckerman, 1952). For example, Tsudzuki (1953) studied the nature of non-response in a two category (yes-no) questionnaire. This was done by administering the same test to the same group with additional categories such as "in-between," "cannot decide," and with two different intensities of "agree" and "disagree." The latter method significantly reduced the percentage of non-response. Ghiselli (1939) noted that the use of yes-no responses generally rated a product's advertising as less sincere than when a four-step scale was used. He felt that people were more willing to respond to a four-step scale. Hughes (1969) concluded that the use of forced choice scales results in a confounding of indifference and awareness.

Efficiency of response, or the ability to use scale points in discriminating among objects and/or attributes, and response style (such as yeasaying) have also been examined in relation to the number of response alternatives. Matell and Jacoby (1972) determined that the proportion of scale used was independent of the number of response alternatives. Several studies suggest that yeasaying tendencies, as measured by the proportion of positive responses, are unaffected by scale length (Goldsamt, 1972; Saunders & Ward, 1964; Tuckman & Lorge, 1953).

The literature reviewed in this section is mostly subject to the limitation of being drawn from college student populations. Subjects of different education, occupational, and age levels may be less predisposed to respond to fine scales.

Scoring and data analysis considerations. Scoring and data analysis considerations may affect the selection of the number of response alternatives to be used in any given study. Several studies compared dichotomous/trichotomous scoring methods to the normal summated scoring procedures and reached the conclusion that results (differences in attitudes) were not affected by the method of scoring (Matell, 1970; Matell and Jacoby, 1971; and Goldsamt, 1972). However, for these specific investigation situations two or three response alternatives might have been the optimal number to employ.

Several problems with a two or three way scoring procedure exist which are statistical in nature. If Chi Square tests are sufficient, two or three categories might be adequate. However, if nonparametric rank order correlations are to be employed, substantial "ties" on ranks will result. Also, if parametric statistics are to be employed, the more alternatives the better, because of the assumption of continuous distributions or interval scale properties. Finally, another analytical issue of concern with the use of two or three point scales is the reproducibility of the original data configuration, an issue important in the use of multidimensional scaling. Using simulated data, Green and Rao (1970) demonstrated that recovery is poor with two or three alternatives, and that diminishing returns set in rapidly beyond eight alternatives. Other considerations related to scoring questionnaires are discussed in Chapter XI.

Summary and Conclusions. The state of the literature was probably best summarized by Ghiselli and Brown in Personnel and Industrial Psychology (1948) and by Guilford in Psychometric Methods (1954). These authorities contend that the optimal number of response alternatives is a matter for empirical determination in any situation, and suggest that there is a wide range of variation in refinement around which the optimal point in reliability changes very little. It would appear, however, that additional research in the area might be warranted covering: the different types of rating scales; various topical areas of research; and subjects with different ability, educational, and sociodemographic characteristics. From such studies more information would be available regarding the optimal number of response alternatives to employ for any specific investigation situation.

Response Anchoring

This section contains a summary of research findings concerning response anchoring, including: types of response anchors; anchored versus unanchored scales; amount of verbal anchoring; selection procedures for verbal scale anchors; and balanced versus unbalanced scales.

Types of response anchors. The researcher's judgment has typically determined whether response anchors are to be verbal, numerical, graphic, or some combination. In its original form, the semantic differential was in the following graphic form (Osgood, Suci & Tannenbaum, 1957). Respondents were instructed to place an X on the line that represented their attitude.

Strong ____: ____: ____: ____: ____: ____: ____: ____Weak

This represents a combination graphic and verbal scale. The Likert method calls for a verbal rating (strongly agree through strongly disagree) to a directional statement phrased either positively or negatively. For example:

> The Modern Volunteer Army (MVA) place too much emphasis on extrinsic factors (e.g., beer in barracks) as opposed to intrinsic, job related factors (e.g., pay, supervision).
> ___Agree Strongly ___Agree ___Undecided ___Disagree ___Disagree Strongly

Few studies were uncovered in the literature review which systematically investigated the effect of type of response alternative employed. In a study to validate consumer attitudes concerning various brands against actual purchases, Abrams (1966) tested four combinations of rating devices:

1. Verbal anchors with a -5 through +5 numerical continuum, e.g.:

Definitely Dislike										Definitely Like
-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5

2. Verbal anchors with a 1 through 10 numerical continuum, e.g.:

Definitely Dislike									Definitely Like
1	2	3	4	5	6	7	8	9	10

3. A verbal and numerical continuum, e.g.:

Dislike Completely	Dislike Somewhat	Dislike a little	Neither like nor Dislike	Like a Little	Like Somewhat	Like Completely
1	2	3	4	5	6	7

4. Verbal continua, e.g.:

Below Average	About Average	A Little Better	A lot Better	One of the Best	None Better

Experimental findings illustrated that: average scale scores are relatively constant regardless of scale type; scale 4 had a lower average prediction error (the differences between predicted brand share and actual consumer purchases); and scale 4 had a far smaller amount of clustering of responses at the extreme positive position. These findings also confirm the conventions of researchers who do not include numerical response alternatives in an attitude measurement scale.

Several additional studies were found which support the use of verbal anchoring and verbally defined response alternatives. In a study employing

Air Force personnel as subjects (Madden, 1964), three forms of rating scales were used: (1) each scale alternative was verbally defined and illustrated; (2) neither definitions nor examples were used (numerical scale); and (3) definitions were used but examples were eliminated (verbal scale). Forms (1) and (3) were equally reliable and of greater reliability than form (2). Form (3) was preferred because it was simpler and less time consuming for raters to use. Peters and McCormick (1966) compared the effectiveness of job-task anchored (verbal) equal appearing intervals scales and simple numerically anchored scales. The job-task anchored scales were found to have significantly greater reliability.

Marsh and Perrin (1925) compared the effectiveness of the graphic scale, percentage scale, and man-to-man scale. On the graphic scale, raters underscored the description most applicable to the subject. On the percentage scale the raters placed a check mark in the column representing the subject's standing, in terms of the perceived amount of a given trait possessed, with reference to a preliminary group of subjects. With the man-to-man scale the subjects were compared with particular individuals representing the standards for the traits rated. The results failed to demonstrate the superiority of any one form of scale, with the range of average deviations from agreement being extremely limited regardless of the form of scale employed. Ross (1966) compared man-to-man job performance ratings with ratings from an anchored rating scale for their validity in guiding salary decisions in a research and development organization. The man-to-man comparison procedure was found to be as valid as the anchored ratings. However, the two methods diverged in important practical ways in the results they produced.

Two other studies reported the favorability of using verbal scale anchors, although neither compared verbal to other types of anchors. In a study of supervisory style of head nurses, Smith and Kendall (1963) anchored evaluative rating scales with examples of expected supervisory behavior. The examples were selected by independent consensus of a number of head nurses. Scale reliabilities ranged above .97.

Barrett, Taylor, Parker and Martins (1958) administered four rating scale formats varying from unstructured to highly structured in nature. Second line supervisors rated clerical workers. The verbal format employing trait titles and behavioral descriptions of scale steps demonstrated higher inter-rater reliability, less halo, and less leniency than did the more structured or less structured formats. It should be noted that this study concerned the amount of verbal cues along a scale.

Based upon the studies reviewed in this section, it appears that empirical support exists to conclude that the reliability of scales with verbal anchors and verbal response alternatives is superior to that of numerical and other combinations of verbal and numerical scales. Little evidence was provided regarding graphic rating techniques. It should be noted that none of the studies addressed the issue of comparative validity or subjects' preferences and/or ability to use the rating instrument.

Anchored versus unanchored scales. A number of studies have been conducted on the topic known as "anchoring effects." This section will focus on differences in research results caused by the use of anchored versus unanchored scales. It should be noted that all the studies which follow compared unanchored scales to scales with one end anchored. Many of the studies varied the anchoring of the left or right end of the scale.

An early study in anchoring effects in the judgment of verbal materials was reported by McGarvey (1943). Subjects were asked to scale statements about the social prestige of occupations and undesirable forms of behavior. Scales used in the experiment were either unanchored or had either one of the two extreme points verbally anchored. Results indicated that in unanchored scales the absolute scale tended to be "stimulus anchored," i.e., anchored by the question stem. With either of the end points anchored, the tendency was to move from the stimulus value on the absolute scale toward the anchored extreme. This anchoring effect has been confirmed by Rogers (1941) and Volkman (1936). Rogers' study also examined confidence in ratings and judgment time. Confidence was only slightly affected due to anchoring, but was found decreasing in higher categories nearer the anchor. Judgment time decreased with anchoring. In a reexamination of Volkman's experiment, Hunt and Volkman (1937) were unable to conclude with certainty that an anchor effect existed. An incomplete shift in scale values occurred. That is, the movement did not include the subject's own stimulus anchor, his most pleasant color.

Several examples of conflicting studies to the above anchor effect investigations are available. For example, Weiss (1961) used two separate experimental groups in a study concerning attitude toward delinquents; the experiments cited above used the same subjects with an anchored scale following an unanchored rating. One group was given an extreme statement as a standard for a punitive category, and the other was given no standard. A contrast effect, movement in ratings away from the anchor statement, was produced by the extreme standard. Hunt (1941) offered evidence supportive of this contrast effect. In conclusion to this experimental study, he commented: "If judgments made with an unanchored scale be repeated with the scale anchored by the further definition of one extreme, there is a shift in the average value of the stimulus judgments, and this shift is in a direction away from the anchoring value." This would mean that when a scale was anchored at its low or negative extreme, the ratings tend to rise or be more positive, and vice versa.

Frawley (1948) had 115 seminarians rate 100 statements about war, belief in God, birth control, etc. His experimental procedure was similar to the previously cited studies -- subjects rated the statements first on an unanchored scale and the next day rated them again on a scale with the most unfavorable end of the scale anchored. The fact that the Spearman rank order correlations were extremely high for the two sets of data indicated minimal presence of anchor effects.

Because of this conflicting evidence it cannot be concluded that use of a single verbal anchor produces anchor effects, contrast effects, or indeed any significant differences in rating scale average scores. Most of the studies seem limited due to the use of the same subjects in simple before (unanchored) and after (anchored) experimental designs. It must also be noted that the investigations cited in this area were theoretical inquiries into common principles or "laws" governing ratings or judgments, and were less concerned with strengths and weaknesses of types of scales. Finally, the fact that not one of the investigations cited employed verbal anchors at both ends of the scale makes it even more difficult to conclude whether anchored or unanchored scales should be employed.

Amount of verbal anchoring. A few research studies have addressed the issue of the effects of varying the amount of verbal anchoring on rating scales. Bendig (1953) examined the impact of amount of verbal anchoring in a study where 225 college students rated themselves as to how much they knew about 12 foreign countries. Alternative scales presented to subjects had the center category defined, the end categories defined, or both center and end categories defined. Results indicated that the reliability of the scales increased with added scale anchoring. In a separate report of the same data base, Bendig and Hughes (1953) concluded that increased verbal anchoring also resulted in a slight increase in the amount of information transmitted by the scale. The use of "information transmitted" here is in the context of information theory, meaning more descriptive data from the respondents.

Abrams (1966) compared seven point scales with only the end categories defined to scales with the entire continuum verbally anchored. The study examined consumer mail panel respondent's attitudes toward national brands of toothpaste and scouring cleanser. Shelf inventories of actual purchases of products in these categories were conducted in a follow-up. Scales which had verbal descriptors for all response alternatives were better predictors of purchase behavior. These scales also displayed greater respondent use of the range of response alternatives, with less clustering at the extreme positive position.

Another study in support of increased verbal anchoring was offered by Madden (1964). Four job evaluation factors were used as the basis of rating 10 Air Force specialties. For each factor three different methods of anchoring were used: (1) each response alternative was defined and illustrated; (2) neither definitions nor examples were used; and (3) definitions were used but examples were omitted. Methods (1) and (3) were approximately equal in reliability, both yielding more reliable scales than Method (2).

Only one empirical study was uncovered which offers evidence somewhat in conflict with the above findings. Carter, Ruggels, and Chaffee (1968) conducted an experiment using 15 semantic differential scales used to describe 12 objects (concepts about schools). One hundred and thirty-five female teachers were given the opportunity to modify the scales during rating. Assuming that not every adjective scale was a useful descriptor, subjects were given one polar adjective for each scale and were asked to fill in the appropriate opposite or note "wouldn't use scale" or "don't know." The authors concluded that for four of the 15 scales, the polar opposite chosen by the subjects was not suggested by Osgood. Also, the authors noted that whatever the merits of anchoring both ends of every scale to measure meaning, it appeàred that subjects can more accurately devote their descriptions to objects when one end of the scale is left for them to describe. Perhaps this study raises more doubts about the validity of the semantic differential technique than it offers concrete evidence regarding the amount of verbal anchoring.

In conclusion, the limited number of research studies cited above are somewhat consistent in reporting greater scale reliability with added verbal anchoring. Also, one experiment (Abrams, 1966) offered evidence of higher predictive validity with more verbal anchoring.

Selection procedures for verbal scale anchors. This section presents literature which dealt with appropriate procedures for the selection of verbal scale anchors. The studies relate to verbal anchors used in Likert scales, semantic differential scales, and rating scales.

A complete bipolar adjective screening methodology for semantic differential scales has been outlined by Lusk (1973). This procedure seems highly applicable to any research considering the use of the semantic differential. The process suggested is as follows:

1. Select from Osgood's Thesaurus Study a set of bipolar adjectives for each factor dimension, evaluative, potency, and activity, applicable to the study.

2. Select a pretest sample representative of the final population in the study.

3. Prepare the pretest concept/adjective test blocks, randomizing the concepts and bipolar adjectives.

4. Administer the pretest semantic differential scales and compare variances (test for differences) from the scale midpoint. The objective here is to eliminate a preponderance of midinterval responses, of each bipolar adjective for each concept evaluated.

5. Order the bipolar adjectives based upon their respective variances, high to low.

6. Select the required number of bipolar adjectives, i.e., those sets with significantly lower variances, F test, from midpoints may be eliminated.

One additional insight into the selection of bipolar adjectives should be mentioned. Carter, Ruggels, and Chaffee (1968) reported that bipolar adjectives such as sweet-ferocious were of little value in rating inanimate objects, such as: "Are boulders sweet-sour?" But they were useful with relational concepts. Again, caution must be exercised in selecting bipolar adjectives or phrases.

Smith and Kendall (1963) tested a procedure for the construction of evaluative rating scales anchored by examples of expected behavior. Expectations, based on having observed similar behavior, were used to permit rating in a variety of situations. Examples, submitted by head nurses as illustrations of nurses' behavior related to a given dimension, were retained only if reallocated to that dimension by other head nurses. They were then scaled as to desirability. Agreement for a number of examples was high, and scale reliabilities ranged above .97.

In general, the interpretation of the above studies is that pretests for the selection of verbal anchors are valuable in building scale content validity and reliability. Rather than employing anchors which *seem* appropriate to the researcher, anchors should preferably be selected by respondents similar to those who will be participating in the study.

Balanced versus unbalanced scales. Historically, the balanced scale has been preferred by researchers. A scale is balanced when it has an equal number of response alternatives on either side of the scale's "indifferent" category. For example, the following verbal scale is balanced:

How would you describe the Volunteer Army?
___ Very Progressive
___ Progressive
___ Moderately Progressive
___ Neither Progressive nor Conservative
___ Moderately Conservative
___ Conservative
___ Very Conservative

Unbalanced scales have been employed when pretest results demonstrated that subjects, by using extreme response alternatives on a scale, produced a skewed distribution of responses rather than the statistically desirable normal distribution around the mean attitude. To minimize "end piling," unbalanced scales have been used. More response alternatives are added to the end of the scale where the piling is likely to occur. This practice tends to shift the distribution of responses along the scale continuum. For example, the following scale is heavily unbalanced on the favorable end:

What is your reaction to the "beer in the barracks" policy?
___ Enthusiastic
___ Extremely Favorable
___ Very Favorable
___ Favorable
___ Fair
___ Poor

Only one empirical study was found in the literature review which dealt with the comparative effects of balanced versus unbalanced scales. In the study (Weiss, 1963b), 350 college students judged the social prestige of 400 occupations. Four types of scales were examined, each with a zero category designating average prestige. They were:

1. Three category balanced with an equal number of plus and minus categories.

2. Seven category balanced with an equal number of plus and minus categories.

3. Five category unbalanced with a single plus category.

4. Five category unbalanced with a single minus category.

The author concluded that relative to the balanced scales, the unbalanced scales induced a shift in the prestige value of the "average" category in the direction of the single-nondiscriminating category. In other words, significant differences between scales occurred. Unfortunately, this investigation did not report data relating to the comparative reliability or validity of balanced and unbalanced scales.

Based upon a single study, obviously no conclusions can be drawn regarding the use of balanced or unbalanced scales. Intuitively, the use of balanced scales seems to be warranted to avoid biasing results with the presence of more favorable (or unfavorable) response categories.

ORDER OF PERCEIVED FAVORABLENESS OF COMMONLY USED WORDS AND PHRASES

This chapter is concerned with the words or phrases which are commonly used as response alternatives in questionnaire items. It is often necessary to arrange the words or phrases along some continuum or in some order of degree, and several studies have been conducted to establish this order. These studies will be discussed in this chapter. Studies concerned with determining the perceived favorableness of words and phrases are described below in terms of the instruments used, type and number of subjects, and method of determining the scale values. When available, lists of words or phrases and scale values have been included.

Major Studies and Lists of Adjectives and Scale Values

One of the first studies on the perceived favorableness of adjectives was conducted by Mosier (1940, 1941a). In this study 296 adjectives were rated on an 11 point scale anchored at 1 with "most unfavorable," at 6 with "neither favorable nor unfavorable," and at 11 with "most favorable." Zero was used if the adjective could not be rated. Each adjective or adjective phrase was judged by approximately 140 students from introductory or second year psychology courses. Twenty-six of the 296 words were scaled by Thurstone's method of successive intervals, using the stimulus "completely unsatisfactory" as the standard, with its mean at zero and its standard deviation equal to one. The medians, scale values, and standard deviations, for these 26 words are given in Table VII-1 (Mosier, 1940). The method of equal appearing intervals was also used to find the scale value for each of the 296 words. A sample list of 14 words is shown in Table VII-2. In this study correlation coefficients were computed on six words (neutral, normal, excellent, desirable, disgusting, and unsatisfactory) which were repeated in the list presented to the subjects. The correlation coefficients for these words ranged from .90 to .99. In Mosier's list there were 26 words that could not be rated by 20 or more subjects. These words are listed in Table VII-3. Some of the words also exhibited marked bimodality of response, and these are shown in Table VII-4. Complete tables showing the results of this study were privately issued by Mosier (1941b), but this list was unavailable for review.

Mosier's research also studied the effect of usual adverbal intensives. A set of five words were selected and these words were repeated with each of seven intensives. The results of the study are given in Table VII-5. Four of the seven adjectives selected are arranged across the top of the table, each heading a column. The fifth adjective, "indifferent," behaved atypically because of ambiguous associated context. Each row of the table

TABLE VII-1

Scale Values of Standard Set of Words, (Mosier, 1940)

Stimulus	Md.	S.V.	S.D.
Completely unsatisfactory	1.5	0.00	1.00
Very unsatisfactory	2.3	0.75	0.65
Catastrophic	2.5	0.91	0.81
Treacherous	2.7	1.05	0.62
Menacing	2.9	1.14	0.56
Discouraging	3.5	1.42	0.49
Painful	3.6	1.43	0.54
Unprofitable	4.3	1.72	0.62
Rejected	4.6	1.79	0.54
Disputable	5.7	2.42	0.69
Normal	6.7	2.47	1.48
Satiating	5.2	2.79	1.64
Reconcilable	6.3	3.20	0.75
Blameless	7.6	3.64	0.90
Solacing	8.0	3.79	0.51
Ordinary	6.5	3.83	1.43
Bonny	8.4	3.97	0.61
Decent	8.5	4.08	0.61
Preferable	9.0	4.30	0.55
Profitable	9.4	4.40	0.47
Popular	9.7	4.55	0.49
Successful	10.0	4.65	0.54
Sublime	10.3	4.90	0.86
Superior	10.4	4.93	0.54
Completely agreeable	10.1	4.95	0.65
Superb	11.1	5.25	0.68

TABLE VII-2

Scale Values of Selected Words (Mosier, 1941a)

Stimulus	Scale Value
Completely unsatisfactory	0.00
Repulsive	0.50
Disgraceful	1.00
Wrong	1.50
Unnecessary	2.00
Disputable	2.36
Excusable	2.83
Average	3.00
Pardonable	3.48
Comfortable	4.04
Desirable	4.50
Highly agreeable	5.02
Divine	5.50
Very, very desirable	5.66

TABLE VII-3

Words Marked "Unable to Rate" by 20 or More Subjects
(Mosier 1941a)

Abhorred
Adverse
Bonny
Calamitous
Cloying
Debased
Despicable

Ecstatic
Estimable
Expedient
Inflaming
Iniquitous
Noxious
Odious

Ominous
Peerless
Pernicious
Persuasive
Perverse
Pestilential

Propitious
Satisfying
Seductive
Seemly
Solacing
Superlative

TABLE VII-4

Words Exhibiting Marked Bimodality of Response
(Mosier 1941a)

Acceptable	Completely indifferent		Irresistible
Amazing	Extremely indifferent	*	Normal
Appalling	Highly indifferent		Peerless
Base	Quite indifferent	*	Satiating
Bearable	Unusually indifferent		Seductive
Bewitching	Very indifferent	*	Sublime
Choice	Very, very indifferent		Tempting
Important	Inflaming		Unfit
Indifferent	Indispensable		Unspeakable

Note. Words marked with asterisks also appear in Table VII-1.

TABLE VII-5

Scale Values as Affected by Adverbial Modifiers
(Mosier 1941a)

Modifier	Desirable	Agreeable	Poor	Unsatisfactory
(Unmodified)	4.50	4.19	1.60	1.47
Quite	4.76	4.45	1.30	1.00
Very	4.96	4.82	1.18	0.75
Unusually	5.23	4.86	0.95	0.75
Completely	5.38	4.96	0.92	0.80
Highly	5.15	5.02		0.71
Extremely	5.42	5.10	0.95	0.10
Very, very	5.66	5.34	0.55	0.25

presents the scale values for one of the adverbial modifiers studied.

Jones and Thurstone (1955), in order to determine the degree of like or dislike denoted by an adjective or phrase, had 905 enlisted personnel rate 51 descriptive words and phrases on a nine point scale anchored with "greatest dislike" at the left, "neither like nor dislike" in the center, and "greatest like" at the right. For each item a scale value was determined by the method of successive intervals and a standard deviation was computed. The 51 word phrases are given in Table VII-6.

Myers and Warner (1968) conducted a study in which 50 commonly used statements describing product taste or ad effectiveness were rated on a 21 point Thurstone equal interval scale with the top category captioned "This is the best thing I could say about the (person, product, or ad)." The bottom and opposite category was "This is the worst thing I could say about the (person, product, or ad)." The judges were 25 housewives, 36 business executives, 40 graduate business administration students, and 25 undergraduate business administration students. For each statement the mean scale values and standard deviations were computed. The 50 statements are given in Table VII-7.

Cliff (1959) reported on a study which derived scale values for 150 evaluative words and phrases. The list of stimuli used 15 unmodified adjectives plus all combinations of them and nine intensity adverbs. Two hundred thirteen students in introductory psychology courses at Wayne State University, 183 at Princeton, and 129 at Dartmouth rated the words and phrases, on an 11 point scale anchored by "most unfavorable" at the left, "neutral" in the center, and "most favorable" at the right. The referent of the items was "favorable or unfavorable opinions about people." Scale values were derived by the least squares, successive interval method. The scale values of the adverb-adjective combinations are shown in Table VII-8. The adverb and adjective values matrices are shown in Table VII-9.

Altemeyer (1970) conducted two studies in which numerical values were assigned adverb-verb combinations. In the first study, 392 introductory psychology students rated eight adverb-verb combinations on a seven point scale with values from minus three to plus three. In the second study, 194 introductory psychology students assigned numerical values to nine adverb-verb combinations on a four point scale ranging from zero to plus three. Plus three was labeled either "completely agree" or "strongly agree." The mean ratings of the verbal phrases obtained for both studies are listed in Table VII-10.

Dodd and Gerberick (1960) presented sets of word phrases to groups of subjects who were to place each item on a nine point scale. For each group of words the median scale position was calculated. Table VII-11 shows the scale positions for 34 phrases rated by 40 subjects. Table VII-12 shows the median scale positions for 47 intensity phrases tested in series context. Table VII-13 shows the findings from 100 judges for

TABLE VII-6

Scale Values and Standard Deviations of Stimulus Items (Jones and Thurstone, 1955)

Item	Scale Value	*SD*
Best of all	6.15	2.48
Favorite	4.68	2.18
Like extremely	4.16	1.62
Like intensely	4.05	1.59
Excellent	3.71	1.01
Wonderful	3.51	.97
Strongly like	2.96	.69
Like very much	2.91	.60
Mighty fine	2.88	.67
Especially good	2.86	.82
Highly favorable	2.81	.66
Like very well	2.60	.78
Very good	2.56	.87
Like quite a bit	2.32	.52
Enjoy	2.21	.86
Preferred	1.98	1.17
Good	1.91	.76
Welcome	1.77	1.18
Tasty	1.76	.92
Pleasing	1.58	.65
Like fairly well	1.51	.59
Like	1.35	.77
Like moderately	1.12	.61
OK	.87	1.24
Average	.86	1.08
Mildly like	.85	.47
Fair	.78	.85
Acceptable	.73	.66
Only fair	.71	.64
Like slightly	.69	.32
Neutral	.02	.18
Like not so well	−.30	1.07
Like not so much	−.41	.94
Dislike slightly	−.59	.27
Mildly dislike	−.74	.35
Not pleasing	−.83	.67
Don't care for it	−1.10	.81
Dislike moderately	−1.20	.41
Poor	−1.55	.87
Dislike	−1.58	.94
Don't like	−1.81	.97
Bad	−2.02	.80
Highly unfavorable	−2.16	1.37
Strongly dislike	−2.37	.53
Dislike very much	−2.49	.64
Very bad	−2.53	.61
Terrible	−3.09	.98
Dislike intensely	−3.33	1.39
Loath	−3.76	3.54
Dislike extremely	−4.32	1.86
Despise	−6.44	3.62

TABLE VII-7

Means and Standard Deviations of Commonly Used Statements
(Myers and Warner, 1968)

Item	*Housewives*		*Executives*		*Graduate business students*		*Undergraduate business students*	
	M	SD	M	SD	M	SD	M	SD
Superior	20.12	(1.17)	18.22	(2.82)	19.45	(1.78)	18.96	(1.67)
Fantastic	20.12	(0.83)	18.69	(3.68)	20.15	(1.37)	19.20	(1.87)
Tremendous	19.84	(1.31)	18.67	(2.01)	19.70	(1.18)	18.92	(1.75)
Superb	19.80	(1.19)	19.00	(2.10)	19.40	(1.95)	19.60	(2.42)
Excellent	19.40	(1.73)	18.72	(2.25)	19.58	(1.97)	19.44	(1.42)
Terrific	19.00	(2.45)	18.81	(2.19)	19.08	(1.61)	18.60	(1.63)
Outstanding	18.96	(1.99)	19.31	(2.01)	19.58	(1.26)	19.40	(1.35)
Exceptionally good	18.56	(2.36)	17.03	(4.12)	17.68	(2.26)	17.88	(1.72)
Extremely good	18.44	(1.61)	17.33	(3.09)	17.45	(2.26)	18.00	(1.50)
Wonderful	17.32	(2.30)	17.97	(2.35)	18.45	(1.99)	17.52	(2.10)
Unusually good	17.08	(2.43)	16.47	(2.99)	16.78	(2.12)	16.20	(1.80)
Remarkably good	16.68	(2.19)	17.44	(2.63)	17.20	(2.32)	17.08	(1.89)
Delightful	16.92	(1.85)	16.61	(2.45)	16.60	(2.24)	16.76	(1.51)
Very good	15.41	(2.77)	16.83	(2.52)	17.00	(2.18)	16.80	(1.44)
Fine	14.80	(2.12)	15.61	(2.72)	14.60	(3.00)	15.32	(2.21)
Quite good	14.44	(2.76)	13.69	(2.90)	15.70	(2.08)	15.60	(1.94)
Good	14.32	(2.08)	13.81	(3.25)	14.78	(2.27)	14.56	(1.96)
Moderately good	13.44	(2.23)	11.42	(2.99)	12.60	(2.55)	13.04	(1.43)
Pleasant	13.44	(2.06)	13.61	(2.43)	13.48	(2.33)	14.48	(2.14)
Reasonably good	12.92	(2.93)	11.89	(3.37)	13.85	(2.19)	14.20	(1.71)
Nice	12.56	(2.14)	11.44	(2.79)	12.70	(2.65)	13.72	(1.77)
Fairly good	11.96	(2.42)	11.94	(3.84)	12.40	(2.24)	13.12	(2.11)
Slightly good	11.84	(2.19)	10.25	(3.14)	11.88	(2.62)	12.32	(1.52)
Acceptable	11.12	(2.59)	10.67	(3.34)	10.72	(1.96)	11.40	(2.02)
Average	10.84	(1.55)	9.97	(2.34)	10.82	(1.43)	10.76	(1.05)
All right	10.76	(1.42)	10.17	(3.28)	10.95	(2.15)	11.40	(1.26)
O.K.	10.28	(1.67)	10.11	(2.48)	10.58	(2.12)	11.28	(1.21)
So so	10.08	(1.87)	8.81	(2.75)	9.52	(1.47)	10.36	(1.15)
Neutral	9.80	(1.50)	9.56	(1.90)	10.18	(2.01)	10.52	(1.16)
Fair	9.52	(2.06)	9.56	(3.67)	9.20	(2.05)	10.24	(2.20)
Mediocre	9.44	(1.80)	8.11	(2.74)	8.90	(2.36)	9.36	(2.20)
Not very good	6.72	(2.82)	6.47	(2.41)	6.40	(2.05)	7.92	(2.02)
Moderately poor	6.44	(1.64)	6.83	(3.50)	6.28	(1.87)	7.24	(1.59)
Reasonably poor	6.32	(2.46)	6.31	(2.19)	5.82	(1.74)	6.16	(1.57)
Slightly poor	5.92	(1.96)	7.19	(2.36)	7.25	(2.00)	8.48	(1.83)
Poor	5.76	(2.09)	5.19	(2.86)	4.72	(2.51)	5.24	(1.31)
Fairly poor	5.61	(1.68)	6.67	(2.81)	6.25	(1.63)	6.72	(1.74)
Unpleasant	5.04	(2.82)	4.36	(3.02)	4.68	(2.63)	5.52	(2.06)
Quite poor	4.80	(1.44)	4.56	(2.58)	3.62	(1.67)	4.56	(1.78)
Bad	3.88	(2.19)	3.67	(2.54)	3.85	(1.81)	4.24	(1.88)
Very bad	3.20	(2.10)	2.22	(2.34)	2.70	(2.16)	3.08	(1.50)
Unusually poor	3.20	(1.41)	3.08	(1.79)	3.48	(1.68)	4.16	(1.57)
Very poor	3.12	(1.17)	3.14	(2.39)	3.35	(1.99)	3.68	(1.52)
Remarkably poor	2.88	(1.71)	2.75	(1.70)	3.12	(1.70)	3.92	(1.68)
Unacceptable	2.64	(2.01)	3.53	(3.42)	3.98	(2.79)	5.56	(3.06)
Exceptionally poor	2.52	(1.19)	3.19	(2.23)	3.22	(1.82)	3.52	(1.96)
Extremely poor	2.08	(1.19)	2.83	(2.14)	3.10	(1.72)	3.24	(1.76)
Awful	1.92	(1.50)	2.25	(1.46)	2.48	(1.72)	2.68	(1.86)
Terrible	1.76	(0.77)	2.22	(2.63)	2.05	(1.43)	1.88	(1.24)
Horrible	1.48	(0.87)	2.22	(2.51)	1.62	(1.15)	2.00	(1.35)

TABLE VII-8

Obtained Successive Intervals Scale Values of Adverb-Adjective Combinations (Cliff, 1959)

Wayne State University

	Contemptible	Immoral	Disgusting	Bad	Inferior	Ordinary	Average	Nice	Good	Pleasant
(Unmodified)	.793	.793	.828	1.024	1.323	2.074	2.145	2.636	2.712	2.770
Slightly	1.324	1.176	1.327	1.497	1.520	1.980	2.023	2.286	2.417	2.440
Somewhat	1.134	1.133	.963	1.323	1.516	2.038	2.080	2.488	2.462	2.505
Rather	1.047	.954	.990	1.232	1.295	2.034	2.172	2.568	2.755	2.743
Pretty	.867	.884	.753	1.018	1.180	2.026	2.094	2.767	2.622	2.738
Quite	.832	.726	.687	.924	1.127	2.023	2.101	2.738	2.880	2.849
Decidedly	.609	.528	.636	.797	1.013	1.949	2.020	2.969	3.024	3.028
Unusually	.661	.451	.664	.662	.963	1.875	2.062	3.155	3.243	3.107
Very	.633	.465	.502	.639	.927	2.073	2.039	3.016	3.250	3.174
Extremely	.327	−.085	.395	.470	.705	1.936	2.052	3.351	3.449	3.490
Princeton University										
(Unmodified)	.858	.956	.836	1.080	1.274	1.869	2.053	2.736	2.910	2.816
Slightly	1.342	1.393	1.273	1.546	1.548	1.928	2.008	2.407	2.449	2.457
Somewhat	1.217	1.232	1.042	1.365	1.486	1.943	2.050	2.511	2.490	2.572
Rather	.995	1.113	.965	1.257	1.396	1.913	2.056	2.685	2.821	2.793
Pretty	1.046	1.034	.822	1.225	1.183	1.902	2.052	2.711	2.727	2.738
Quite	.884	.893	.736	1.012	.974	1.871	1.951	2.778	2.990	2.920
Decidedly	.742	.791	.718	.889	.986	1.763	1.828	3.009	3.095	3.033
Unusually	.753	.701	.652	.808	.926	1.660	1.877	3.206	3.304	3.240
Very	.744	.765	.593	.834	.899	1.785	1.872	3.042	3.250	3.172
Extremely	.556	.494	.473	.666	.758	1.806	1.800	3.282	3.443	3.437
Dartmouth College										
(Unmodified)	.801	1.002	.792	1.125	1.329	2.043	1.984	2.669	2.770	2.830
Slightly	1.336	1.328	1.270	1.466	1.487	2.005	2.052	2.366	2.411	2.402
Somewhat	1.148	1.227	1.056	1.343	1.452	2.006	2.003	2.483	2.521	2.601
Rather	1.032	1.099	.935	1.266	1.347	1.964	2.048	2.594	2.778	2.811
Pretty	1.015	.974	.749	1.065	1.123	2.025	2.044	2.739	2.686	2.726
Quite	.755	.839	.661	1.004	.985	1.920	1.950	2.760	2.946	2.887
Decidedly	.647	.754	.625	.920	.970	1.969	1.864	2.968	3.052	3.034
Unusually	.663	.653	.571	.837	.893	1.824	1.989	3.170	3.262	3.123
Very	.652	.690	.538	.828	.900	1.903	1.977	3.080	3.228	3.103
Extremely	.465	.372	.325	.725	.703	1.790	1.799	3.326	3.410	3.465

TABLE VII-3

Adverb and Adjective Value Matrices
(Cliff, 1959)

	Wayne		Princeton		Dartmouth	
Adverb	Actual Value	Expected Value	Actual Value	Expected Value	Actual Value	Expected Value
(Unmodified)	1.000	.987	1.000	.993	1.000	.991
Slightly	.555	1.000	.559	.999	.538	1.003
Somewhat	.685	.997	.719	1.001	.662	.995
Rather	.846	1.015	.887	1.014	.843	1.016
Pretty	.935	.995	.961	.994	.878	.992
Quite	1.042	.994	1.109	.988	1.047	.991
Decidedly	1.216	.997	1.231	.996	1.165	.992
Unusually	1.291	1.010	1.324	1.001	1.281	1.010
Very	1.317	1.008	1.323	1.007	1.254	1.002
Extremely	1.593	.996	1.546	.997	1.446	1.006
Adjective						
Evil	-1.246	2.082	-.989	1.918	-.993	1.972
Wicked	-1.158	1.952	-.951	1.848	-.997	1.910
Contemptible	-.913	1.746	-.826	1.749	-.882	1.792
Immoral	-1.177	1.936	-.931	1.878	-.954	1.910
Disgusting	-.806	1.617	-.801	1.621	-.902	1.715
Bad	-1.025	2.032	-.972	2.051	-.796	1.907
Inferior	-.813	2.008	-.923	2.077	-.861	2.037
Ordinary	-.078	2.083	-.253	2.100	-.223	2.182
Average	-.040	2.121	-.296	2.254	-.211	2.195
Nice	1.007	1.742	.984	1.842	1.011	1.739
Good	1.078	1.752	1.158	1.777	1.075	1.761
Pleasant	1.001	1.835	1.050	1.856	.974	1.860
Charming	.802	2.136	.895	2.116	.910	2.013
Admirable	.983	2.001	1.170	1.892	1.086	1.892
Lovable	.836	2.173	.912	2.108	.812	2.207

TABLE VII-10

Numerical Ratings of Adverb-Verb Combinations
(Altemeyer, 1970)

Adverb	Study 1				Study 2	
	Disagree		Agree		Agree	
	M	*SD*	*M*	*SD*	*M*	*SD*
Slightly	-.64	.38	.67	.36	0.62	.31
Substantially	-2.17	.51	2.10	.50	2.08	.49
Moderately	-1.35	.42	1.47	.41	1.49	.38
Somewhat	-.93	.47	.94	.41	.91	.42
Quite	-2.16	.57	2.37	.49	2.23	.46
Considerably	-2.17	.45	2.21	.42	2.18	.40
Perhaps	-.43	.46	.52	.46	.44	.43
Decidedly	-2.76	.43	2.77	.41	2.74	.47
Mildly					.98	.41

TABLE VII-11

Scale Positions for Thirty-four Phrases (Dodd and Gerberick, 1960)

	Median
Degree phrases, tested out-of-context	
complete	8.85
almost complete	8.06
very much more	8.02
much more	7.67
a lot more	7.50
a good deal more	7.29
more	6.33
somewhat more	6.25
a little more	6.00
slightly more	5.99
now	5.03
AS AT PRESENT	5.00
slightly less	3.97
a little less	3.96
somewhat	3.79
less	3.64
much less	2.55
a good deal less	2.44
a lot less	2.36
very little	2.08
almost none	2.04
very much less	1.96
none	1.11
Temporal frequency phrases, tested out-of-context	
always	8.99
without fail	8.89
often	7.23
usually	7.17
frequently	6.92
now and then	4.79
sometimes	4.78
occasionally	4.13
seldom	2.45
rarely	2.08
never	1.00

TABLE VII-12

Scale Positions of 47 Intensity Phrases
(Dodd and Gerberick, 1960)

	Median
Subsets of intensity phrases, tested in series context	
Subset 1	
very strongly	8.40
strongly	7.07
moderately	5.24
indifferent	3.70
Subset 2	
very certain	8.55
certain	7.13
uncertain	2.83
not certain	2.64
Subset 3	
extremely vital	8.79
very vital	7.55
vital	5.92
insignificant	2.12
Subset 4	
of great importance	8.05
don't care	4.63
doesn't mean anything	2.50
Subset 5	
very sure	8.15
sure	5.93
not sure	2.82
Subset 6	
without hesitation	7.50
with little hesitation	5.83
hesitant	4.77
with some hesitation	4.38
with considerable hesitation	3.29
with much hesitation	3.20
with great hesitation	2.41
only as a last resort	1.70
Subset 7	
very good	8.08
good	6.72
fair	4.96
bad	2.83
very bad	1.50
Subset 8	
essential	7.58
non-essential	2.58
Subset 9	
very urgent	8.00
urgent	6.41
Subset 10	
very important	8.22
important	6.83
don't know	4.82
undecided	4.76
not important	3.09
unimportant	2.94
very unimportant	1.75
Subset 11	
object strongly to	3.50
feel strongly toward	7.80
Subset 12	
very crucial	8.29
crucial	6.39
doesn't make any difference	2.83

TABLE VII-13

Stability of Intensity Phrases in Diverse Contexts
(Dodd and Gerberick, 1960)

Intensity Phrase	Issue	N number of responses	Scale position	Mean scale position
Very strongly	1	295	8.96	8.92
	2	197	8.91	
	3	161	8.91	
Strongly	1	269	7.01	7.11
	2	162	7.20	
	3	271	7.12	
Moderately	1	305	4.78	4.82
	2	189	4.77	
	3	277	4.92	
Indifferent		(Insufficient data)		

the phrases of Subset 1 of Table VII-12 on strength of feeling, when presented in graded series and as applied to 31 scale statements about three issues. The three issues, respectively, were: resistance to starting a war; drafting of women for military service and defense work; and amount of government control. As seen in the table, Dodd and Gerberick found that the diversity of context does not appreciably shift the scores of the intensity phrases.

In a study conducted by Anderson (1968) a sample of 100 college students rated 555 personality trait words on likableness as a personality characteristic. The words were on a seven point scale with zero being defined as "least favorable or desirable" and 6 as "most favorable or desirable." The words were also rated for meaningfulness by 50 subjects on a scale that ranged from zero ("I have almost no idea of the meaning of this word") to 4 ("I have a very clear and definite understanding of the meaning of this word"). Table VII-14 shows the list of words in order of likableness. The first entry for each word is its likableness value, listed in the column headed "L". The L value is the sum of the ratings of the 100 subjects so the mean may be obtained by inserting a decimal point. The second entry of the table in the column headed s^2 is the variance of the likableness ratings.

In a recent study (Matthews, Wright, & Yudowitch, 1975) a list of 141 adjective phrases showing degrees of adequacy, acceptability, and comparison were administered to enlisted men and officers at Fort Hood. The adjective phrases were rated on an eleven point scale with -5 anchored with "most unfavorable", zero anchored with "neither unfavorable nor favorable", and +5 with "most favorable". Means and standard deviations were computed for each adjective phrase. There were about 50 usable judgements for each phrase. The results of this study are shown in Tables VII-15, VII-16, and VII-17.

Currently, The U.S. Army Test and Evaluation Command (1973) is carrying out a project part of which included the scaling of 32 adjectives and adjective phrases. Average scale scores and standard deviations were computed for the list of adjectives. The 32 adjectives and adjective phrases are shown in Table VII-18.

Simpson (1944) studied the commonly held meaning of 20 words denoting frequency by having 335 high school and college students respond to how many items out of 100 each word in a list indicated. The results are presented in Table VII-19.

A significant study conducted by Mittelstaedt (1971) compared the results of the Jones and Thurstone (1955), Cliff (1959), and Myers and Warner (1968) studies. The Jones and Thurstone and the Myers and Warner studies had 13 stimuli in common. Values for the same 13 stimuli (treating them as a "scale") were taken from each of the Myers and Warner groups (housewives, executives, graduate students, and undergraduates). Product moment correlation coefficients between the Jones and Thurstone "scale" and the values for each of the Myers and Warner groups were then calculated. The results are shown in Table VII-20. Eleven stimuli in Cliff's study were the same as 11 in the Myers and Warner study. As before the 11 items for each of Cliff's study groups were treated as a "scale" and were compared with a "scale" constructed using the same stimuli for each of the four Myers and Warner subject groups. The product moment correlation coefficients for each Cliff group with each Myers and Warner group are presented in Table VII-21. As may be seen in the tables, Mittelstaedt found a remarkable correspondence among the scale values from the three studies in spite of differences in time, place, subjects, instrumentation, instructions, referents, and context.

TABLE VII-14

Ratings of Likableness, and Likableness Variances for 555 Common Personality Traits Arranged in Order of Decreasing Likableness (Anderson, 1968)

Word	L	s^2	Word	L	s^2
sincere	573	.30	conscientious	481	.82
honest	555	.47	resourceful	481	.74
understanding	549	.52	alert	480	.65
loyal	547	.60	good	480	.99
truthful	545	.61	witty	480	.81
trustworthy	539	.62	clear-headed	479	.69
intelligent	537	.62	kindly	479	1.06
dependable	536	.66	admirable	478	.78
open-minded	530	.56	patient	478	.70
thoughtful	529	.47	talented	478	.84
wise	528	.61	perceptive	477	.84
considerate	527	.76	spirited	477	.64
good-natured	527	.82	sportsmanlike	477	1.11
reliable	527	.66	well-mannered	477	1.05
mature	522	.66	cooperative	476	.85
warm	522	.60	ethical	476	1.15
earnest	521	.73	intellectual	476	.91
kind	520	.69	versatile	474	.66
friendly	519	.72	capable	471	.63
kind-hearted	514	.87	courageous	471	.85
happy	514	.77	constructive	468	.46
clean	514	.99	productive	468	.81
interesting	511	.64	progressive	468	.78
unselfish	510	.68	individualistic	467	1.50
good-humored	507	.73	observant	467	.81
honorable	507	.85	ingenious	466	.75
humorous	505	.86	lively	466	.75
responsible	505	.76	neat	466	.93
cheerful	504	.83	punctual	466	1.26
trustful	504	1.07	logical	465	.76
warm-hearted	504	.62	prompt	465	1.16
broad-minded	503	.80	accurate	464	.98
gentle	503	1.00	sensible	464	.84
well-spoken	501	.78	creative	462	1.15
educated	500	.73	self-reliant	462	.96
reasonable	500	.73	tolerant	461	.91
companionable	499	.88	amusing	460	.89
likable	497	.78	clean-cut	460	1.49
trusting	497	1.20	generous	459	.89
clever	496	.56	sympathetic	459	1.05
pleasant	495	.86	energetic	457	.81
courteous	494	.94	high-spirited	457	.73
quick-witted	494	.78	self-controlled	456	.69
tactful	494	.84	tender	456	1.30
helpful	492	.74	active	455	.65
appreciative	492	.78	independent	455	1.32
imaginative	492	.96	respectable	455	1.10
outstanding	492	1.00	inventive	453	.86
self-disciplined	491	.75	wholesome	453	1.14
brilliant	490	.96	congenial	452	.82
enthusiastic	489	.72	cordial	452	.96
level-headed	489	.68	experienced	451	.76
polite	489	1.11	attentive	450	.84
original	488	.75	cultured	450	.80
smart	488	.65	frank	450	1.10
forgiving	486	1.03	purposeful	450	.86
sharp-witted	486	1.01	decent	449	1.00
well-read	486	.67	diligent	449	.82
ambitious	484	1.14	realist	449	.94
bright	483	.67	eager	448	.80
respectful	483	1.17	poised	448	.78
efficient	482	.94	competent	447	.82
good-tempered	482	1.02	realistic	447	.90
grateful	482	1.00	amiable	446	1.02

(Table continued on next page)

Ratings of Likableness, and Likableness Variances for 555 Common Personality Traits Arranged in Order of Decreasing Likableness (Anderson, 1968)

Word	L	s^2	Word	L	s^2
optimistic	443	1.30	soft-hearted	387	1.69
vigorous	443	.81	dignified	386	1.05
entertaining	442	.63	philosophical	386	1.78
adventurous	441	.90	idealistic	384	1.35
vivacious	440	.91	soft-spoken	380	1.03
composed	439	.87	disciplined	379	1.24
relaxed	439	.99	serious	379	.89
romantic	439	1.19	definite	375	.76
proficient	438	.70	convincing	374	.76
rational	438	1.37	persuasive	374	.92
skillful	438	.80	obedient	373	1.67
enterprising	437	.76	quick	373	1.33
gracious	437	1.04	sophisticated	372	.95
able	436	.68	thrifty	372	.75
nice	436	1.28	sentimental	371	1.10
agreeable	434	.95	objective	370	1.81
skilled	433	.83	nonconforming	369	1.33
curious	432	1.13	righteous	369	2.24
modern	432	.93	mathematical	367	1.01
charming	430	.98	meditative	366	1.52
sociable	429	.85	fearless	366	1.12
modest	428	1.25	systematic	366	1.12
decisive	427	1.03	subtle	365	1.00
humble	427	1.51	normal	362	1.21
tidy	427	.82	daring	360	1.03
popular	426	.98	middleclass	360	.99
upright	426	1.04	lucky	358	1.30
literary	425	1.46	proud	358	1.66
practical	425	.73	sensitive	358	2.00
light-hearted	424	.99	moralistic	357	2.13
well-bred	423	1.13	talkative	352	1.32
refined	422	1.16	excited	351	.86
self-confident	421	.81	moderate	351	.90
cool-headed	420	.97	satirical	351	1.18
studious	418	1.00	prudent	348	1.71
venturesome	417	.85	reserved	348	1.00
discreet	416	1.29	persistent	347	1.66
informal	416	1.00	meticulous	346	1.38
thorough	416	.94	unconventional	346	.92
exuberant	414	.97	deliberate	345	1.40
inquisitive	413	1.47	painstaking	345	1.44
easygoing	412	1.20	bold	336	1.22
outgoing	412	1.46	suave	335	1.40
self-sufficient	412	1.30	cautious	334	.77
casual	411	1.11	innocent	332	1.27
consistent	411	1.01	inoffensive	332	.91
moral	411	1.67	shrewd	328	2.47
self-assured	411	.72	methodical	325	1.54
untiring	410	.98	nonchalant	324	1.23
hopeful	406	.92	self-contented	324	2.04
calm	406	.84	perfectionistic	322	1.69
strong-minded	404	1.27	forward	318	1.12
positive	403	1.28	excitable	317	1.15
confident	401	1.04	outspoken	313	1.77
artistic	400	1.58	prideful	313	1.99
precise	400	1.05	quiet	311	.91
scientific	400	1.05	impulsive	307	1.58
orderly	399	.84	aggressive	304	1.43
social	398	1.05	changeable	297	1.08
direct	396	1.07	conservative	295	.92
careful	390	.84	shy	291	.89
candid	389	1.43	hesitant	290	.76
comical	389	1.09	unpredictable	290	1.26
obliging	389	1.53	solemn	289	.85
self-critical	389	1.55	blunt	287	1.63
fashionable	387	1.28	self-righteous	287	2.46
religious	387	1.93	average	284	.90

(Table continued on next page)

TABLE VII-14 (cont.)

Ratings of Likableness, and Likableness Variances for 555 Common Personality Traits Arranged in Order of Decreasing Likableness (Anderson, 1968)

Word	L	s^2	Word	L	s^2
discriminating	283	3.48	spendthrift	221	.73
emotional	283	1.23	temperamental	221	1.10
unlucky	280	.52	gullible	219	.88
bashful	279	.65	indecisive	219	.90
self-concerned	279	1.64	silly	219	1.53
authoritative	274	1.81	submissive	219	.90
lonesome	274	1.06	unstudious	218	1.06
restless	274	.76	preoccupied	216	1.12
choosy	272	1.62	tense	215	.90
self-possesssed	272	2.53	fearful	214	.69
naive	270	1.06	unromantic	214	1.33
opportunist	270	2.47	absent-minded	213	1.00
theatrical	269	1.59	impractical	213	1.12
unsophisticated	267	1.23	withdrawn	213	.80
impressionable	266	.91	unadventurous	212	.93
ordinary	266	.77	sarcastic	210	1.30
strict	266	1.30	sad	209	.93
skeptical	264	1.52	unemot[illegible]	209	1.50
extravagant	263	.88	worrying	209	.71
forceful	263	1.65	high-strung	208	1.57
cunning	262	2.18	unoriginal	207	.81
inexperienced	262	.66	unpoised	206	.76
unmethodical	262	.86	compulsive	205	1.20
daredevil	261	1.23	worrier	205	1.00
wordy	261	1.05	demanding	203	.94
daydreamer	260	.95	unhappy	203	.98
conventional	260	.95	indifferent	202	1.31
materialistic	260	1.66	uncultured	201	1.00
self-satisfied	260	2.00	clumsy	199	.92
rebellious	258	1.40	insecure	198	.75
eccentric	257	1.58	unentertaining	198	.65
opinionated	257	1.98	imitative	198	1.17
stern	257	1.10	melancholy	198	1.13
lonely	256	1.02	mediocre	197	1.10
dependent	254	1.97	obstinate	197	.94
unsystematic	253	.92	unhealthy	197	1.42
self-conscious	249	.92	headstrong	196	1.17
undecided	249	.86	nervous	196	.83
resigned	248	1.22	nonconfident	196	.87
clownish	247	1.73	stubborn	196	1.31
anxious	246	.90	unimaginative	195	1.06
conforming	246	1.26	down-hearted	194	.97
critical	243	1.46	unobservant	194	.90
conformist	241	1.15	inconsistent	193	.91
radical	241	1.80	unpunctual	192	.96
dissatisfied	239	1.65	unindustrious	191	.81
old-fashioned	239	1.39	disturbed	189	.97
meek	238	1.37	superstitious	189	1.33
frivolous	237	1.55	frustrated	188	.93
discontented	237	1.00	illogical	186	.97
troubled	235	.71	rash	186	.59
irreligious	234	1.74	unenthusiastic	186	1.05
overcautious	229	.55	inaccurate	185	.59
silent	228	.83	noninquisitive	184	.90
tough	228	1.74	unagreeable	184	1.08
ungraceful	228	.87	jumpy	183	.73
argumentative	227	1.25	possessive	183	1.62
withdrawing	227	.78	purposeless	183	1.90
uninquisitive	225	.94	moody	182	1.36
forgetful	224	.83	unenterprising	180	.81
inhibited	224	.87	unintellectual	180	1.17
unskilled	224	.71	unwise	180	.79
crafty	223	1.98	oversensitive	179	.77
passive	223	.97	inefficient	178	.68
immodest	222	1.61	reckless	178	1.42
unpopular	222	.80	pompous	177	1.43
timid	222	.78	uncongenial	175	.69

(Table continued on next page)

TABLE VII-14 (cont.)

Ratings of Likableness, and Likableness Variances for 555 Common Personality Traits Arranged in Order of Decreasing Likableness (Anderson, 1968)

Word	L	s^2	Word	L	s^2
untidy	175	.92	tiresome	130	.70
unaccomodating	174	.68	disobedient	128	1.23
noisy	173	.88	complaining	127	.74
squeamish	172	.97	lifeless	127	.68
cynical	171	1.26	vain	127	.99
angry	169	.90	lazy	126	.88
listless	169	.72	unappreciative	126	.84
uninspiring	169	.61	maladjusted	123	1.07
unintelligent	168	1.07	aimless	122	1.16
domineering	167	1.52	boastful	122	.74
scolding	166	.67	dull	121	.81
depressed	166	1.01	gossipy	119	.96
unobliging	165	.86	unappealing	119	1.04
pessimistic	164	1.06	hypochondriac	118	.88
unattentive	164	.74	irritating	118	.67
boisterous	163	1.10	petty	118	.73
suspicious	163	.88	shallow	118	1.00
inattentive	162	1.13	deceptive	117	1.01
overconfident	162	.88	grouchy	117	.61
smug	161	.68	egotistical	116	1.25
unsociable	161	1.13	meddlesome	116	.62
unproductive	160	.65	uncivil	116	.96
wasteful	160	.67	cold	113	.94
fickle	159	1.13	unsportsmanlike	113	.72
neglectful	159	.59	bossy	112	.89
short-tempered	159	.85	unpleasing	112	.71
hot-headed	158	1.09	cowardly	110	.82
unsocial	158	1.16	discourteous	110	.80
envious	157	.77	incompetent	110	.68
overcritical	157	.85	childish	109	.81
scheming	156	1.50	superficial	109	.95
sly	156	1.58	ungrateful	109	.71
weak	155	1.02	self-conceited	108	1.14
foolhardy	154	1.00	hard-hearted	107	1.00
immature	154	.88	unfair	107	1.00
dominating	153	1.28	irresponsible	106	1.17
showy	153	.92	prejudiced	106	1.33
sloppy	153	.96	bragging	104	.72
unsympathetic	153	1.32	jealous	104	.77
uncompromising	153	1.26	unpleasant	104	.81
hot-tempered	152	1.06	unreliable	104	.93
neurotic	152	1.34	impolite	103	.72
unsporting	152	.80	crude	102	1.29
finicky	150	.68	nosey	102	.67
resentful	150	.90	humorless	101	.82
unruly	150	.88	quarrelsome	101	.72
fault-finding	148	.96	abusive	100	.83
messy	147	.78	distrustful	99	1.24
misfit	147	1.28	intolerant	98	.97
uninteresting	146	.78	unforgiving	98	.71
scornful	145	.83	boring	97	.76
antisocial	144	1.24	unethical	97	.90
irritable	143	.85	unreasonable	97	.86
stingy	143	.69	self-centered	96	1.13
tactless	142	.85	snobbish	96	.87
careless	140	.91	unkindly	96	.64
foolish	140	.83	ill-mannered	95	.76
troublesome	140	.73	ill-tempered	95	.62
ungracious	140	.71	unfriendly	92	.80
negligent	139	.68	hostile	91	.77
wishy-washy	139	1.17	dislikable	90	.78
profane	137	1.65	ultra-critical	90	.98
gloomy	136	.84	offensive	88	.83
helpless	136	1.12	belligerent	86	.79
disagreeable	134	.67	underhanded	86	1.19
touchy	134	.83	annoying	84	.66
irrational	130	.70	disrespectful	83	.79

(Table continued on next page)

TABLE VII-14 (cont.)

Ratings of Likableness, and Likableness Variances for 555 Common Personality Traits Arranged in Order of Decreasing Likableness (Anderson, 1968)

Word	L	s^2	Word	L	s^2
loud-mouthed	83	.87	unkind	66	.71
selfish	82	.65	untrustworthy	65	.63
narrow-minded	80	.58	deceitful	62	.96
vulgar	79	1.10	dishonorable	52	.47
heartless	78	.92	malicious	52	.49
insolent	78	.88	obnoxious	48	.60
thoughtless	77	.76	untruthful	43	.43
rude	76	.79	dishonest	41	.51
conceited	74	.84	cruel	40	.54
greedy	72	.61	mean	37	.48
spiteful	72	.61	phony	27	.30
insulting	69	.86	liar	26	.36
insincere	66	.65			

TABLE VII-15

Means and Standard Deviations for Phrases of Degrees of Adequacy
(Matthews, Wright, and Yudowitch, 1975)

Phrase	Mean	SD
Totally adequate	4.620	.846
Absolutely adequate	4.540	.921
Completely adequate	4.490	.825
Extremely adequate	4.412	.719
Exceptionally adequate	4.380	.869
Entirely adequate	4.340	.863
Wholly adequate	4.314	1.038
Fully adequate	4.294	.914
Very very adequate	4.063	.876
Perfectly adequate	3.922	1.026
Highly adequate	3.843	.606
Most adequate	3.843	.978
Very adequate	3.420	.851
Decidedly adequate	3.140	1.536
Considerably adequate	3.020	.874
Quite adequate	2.980	.979
Largely adequate	2.863	.991
Substantially adequate	2.608	1.030
Reasonably adequate	2.412	.771
Pretty adequate	2.306	.862
Rather adequate	1.755	.893
Mildly adequate	1.571	.670
Somewhat adequate	1.327	.793
Slightly adequate	1.200	.566
Barely adequate	.627	.928
Neutral	.000	.000
Borderline	-.020	.316
Barely inadequate	-1.157	.638
Mildly inadequate	-1.353	.621
Slightly inadequate	-1.380	.772
Somewhat inadequate	-1.882	.732
Rather inadequate	-2.102	.974
Moderately inadequate	-2.157	1.017
Fairly inadequate	-2.216	.800
Pretty inadequate	-2.347	.959
Considerably inadequate	-3.600	.680
Very inadequate	-3.735	.777
Decidedly inadequate	-3.780	.944
Most inadequate	-3.980	1.545
Highly inadequate	-4.196	.741

(Table continued on next page)

TABLE VII-15 (Cont.)

Means and Standard Deviations for Phrases of
Degrees of Adequacy
(Matthews, Wright, and Yudowitch, 1975)

Phrase	Mean	SD
Very very inadequate	-4.460	.537
Exceptionally inadequate*	-4.560	.637
Extremely inadequate	-4.608	.527
Fully inadequate	-4.667	.676
Exceptionally inadequate	-4.680	.508
Wholly inadequate	-4.784	.498
Entirely inadequate	-4.792	.644
Completely inadequate	-4.800	.529
Absolutely inadequate	-4.880	.431
Totally inadequate	-4.900	.412

Note. * Indicates duplicated phrase.

TABLE VII-16

Means and Standard Deviations for Phrases of Degrees of Acceptability
(Matthews, Wright, Yudowitch, 1975)

Phrase	Mean	SD
Wholly acceptable	4.725	.563
Completely acceptable	4.686	.610
Fully acceptable	4.412	.867
Extremely acceptable	4.392	.716
Most acceptable	4.157	.915
Very very acceptable	4.157	.825
Highly acceptable	4.040	.631
Quite acceptable	3.216	.956
Largely acceptable	3.137	.991
Acceptable	2.392	1.456
Reasonably acceptable	2.294	.722
Moderately acceptable	2.280	.722
Pretty acceptable	2.000	1.125
Rather acceptable	1.939	.818
Fairly acceptable	1.840	.924
Mildly acceptable*	1.804	.950
Mildly acceptable*	1.686	.700
Somewhat acceptable	1.458	1.241
Barely acceptable	1.078	.518
Slightly acceptable	1.039	.522
Sort of acceptable	.940	.645
Borderline	.000	.200
Neutral	.000	.000
Marginal	-.120	.515
Barely unacceptable	-1.100	.300
Slightly unacceptable	-1.255	.589
Somewhat unacceptable	-1.765	.674
Rather unacceptable	-2.020	.836
Fairly unacceptable	-2.160	.880
Moderately unacceptable	-2.340	.681
Pretty unacceptable	-2.412	.662
Reasonably unacceptable	-2.440	.753
Unacceptable	-2.667	1.381
Substantially unacceptable	-3.235	.899
Quite unacceptable	-3.388	1.066
Largely unacceptable	-3.392	.818
Considerably unacceptable	-3.440	.779
Notably unacceptable	-3.500	1.044
Decidedly unacceptable	-3.837	1.017
Highly unacceptable*	-4.220	.576
Highly unacceptable*	-4.294	.535

(Table continued on next page)

TABLE VII-16 (Cont.)

Means and Standard Deviations for Phrases of
Degrees of Acceptability
(Matthews, Wright, Yudowitch, 1975)

Phrase	Mean	SD
Most unacceptable	-4.420	.724
Very very unacceptable	-4.490	.500
Exceptionally unacceptable	-4.540	.607
Extremely unacceptable	-4.686	.464
Completely unacceptable	-4.900	.361
Entirely unacceptable	-4.900	.361
Wholly unacceptable	-4.922	.269
Absolutely unacceptable	-4.922	.334
Totally unacceptable	-4.941	.235

Note. * Indicates duplicated phrases.

TABLE VII-17

Means and Standard Deviations for Phrases
Used for Comparison
(Matthews, Wright, and Yudowitch, 1975)

Phrase	Mean	SD
Best of all	4.896	.510
Absolutely best	4.843	.459
Truly best	4.600	.721
Undoubtedly best	4.569	.823
Decidedly best	4.373	.839
Best	4.216	1.459
Absolutely better	4.060	.988
Extremely better	3.922	.882
Substantially best	3.700	.922
Decidedly better	3.412	.933
Conspicuously better	3.059	.802
Moderately better	2.255	.737
Somewhat better	1.843	.801
Rather better	1.816	.719
Slightly better	1.157	.776
Barely better	.961	.656
Absolutely alike	.588	1.623
Alike	.216	.847
The same	.157	.801
Neutral	.000	.000
Borderline	-.061	.314
Marginal	-.184	.919
Barely worse	-1.039	.816
Slightly worse	-1.216	.498
Somewhat worse	-2.078	.860
Moderately worse	-2.220	.944
Noticeably worse	-2.529	1.036
Worse	-2.667	1.423
Notably worse	-3.020	1.038
Largely worse	-3.216	1.108
Considerably worse	-3.275	1.206
Conspicuously worse	-3.275	.887
Much worse	-3.286	.808
Substantially worse	-3.460	.899
Decidedly worse	-3.760	.907
Very much worse	-3.941	.752
Absolutely worse	-4.431	.823
Decidedly worst	-4.431	.748
Undoubtedly worst	-4.510	.872
Absolutely worst	-4.686	1.291
Worst of all	-4.776	1.298

TABLE VII-18

Scale Scores of Statements
Based on Over-All Acceptability
(USA,TECOM, 1973)

Statement	Average	Standard Deviation
Excellent	6.27	0.54
Perfect in every respect	6.22	0.86
Extremely good	5.74	0.81
Very good	5.19	0.75
Unusually good	5.03	0.98
Very good in most respects	4.62	0.72
Above average	4.56	0.75
Quite satisfactory	4.35	0.95
Good	4.25	0.90
More than adequate	4.13	1.11
About average	3.77	0.85
Satisfactory	3.69	0.87
Moderately good	3.58	0.77
Adequate	3.39	0.87
Could use some minor changes	3.28	1.09
Not good enough for extreme conditions	3.10	1.30
Not good for rough use	2.72	1.15
Not quite adequate	2.40	0.85
Not very satisfactory	2.11	0.76
Barely adequate	2.10	0.84
Not very good	2.10	0.85
Below average	2.03	0.79
Unsatisfactory but usable	2.00	0.87
Needs major changes	1.97	1.12
Not adequate	1.83	0.98
Barely acceptable	1.79	0.90
Not good enough for general use	1.76	1.21
Better than nothing	1.22	1.08
Poor	1.06	1.11
Very poor	0.76	0.95
Very unsatisfactory	0.69	1.32
Extremely poor	0.36	0.76

TABLE VII-19

Meaning of Frequency Words
(Simpson, 1944)

Term	75% of Students Thought the Term Meant Less Than This Percentage of the Time
Always	100
Very often	93
Usually	90
Often	85
Generally	85
Frequently	80
Rather often	80
About as often as not	50
Now and then	35
Sometimes	35
Occasionally	33
Once in a while	27
Not often	20
Seldom	18
Usually not	18
Hardly ever	13
Very seldom	10
Rarely	10
Almost never	5
Never	2

TABLE VII-20

Correlations of Jones and Thurstone and Myers and Warner "Scale" Values for 13 Stimuli (Mittelstaedt, 1971)

Myers-Warner Groups	r
Housewives	.992
Executives	.986
Graduates	.989
Undergraduates	.993

TABLE VII-21

Correlations of Myers-Warner and Cliff Scale Values for 11 Stimuli (Mittelstaedt, 1971)

Myers-Warner Groups	Cliff Study Groups		
	Wayne State	Princeton	Dartmouth
Housewives	.990	.990	.987
Executives	.990	.988	.989
Graduates	.993	.994	.991
Undergraduates	.996	.997	.995

TABLE VII-22

Summary of Studies on Perceived Favorableness of Commonly Used Words and Phrases

Experimenter	Type of Subjects	No. of Subjects	Type of Words	No. of Words
Mosier (1940, 1941a, 1941b)	Psychology students	140	Adjectives	289
Jones & Thurstone (1955)	Army enlisted personnel	905	Adverbs Adjectives	7 51
Myers & Warner (1968)	Housewives, Business executives, Graduate business administration students, Undergraduate business administration students	25 36 40 25	Adjectives	50
Cliff (1959)	Undergraduate students	537	Adverbs	9
Altemeyer (1970)	College students	586	Adverbs	8
Dodd & Gerberick (1960)	Unknown	40	Adjectives	81
Anderson (1968)	College students	100	Personality traits words	555
USA, TECOM (1973)	Unknown	Unk.	Adjectives	32
Simpson (1944)	High school and college students	100	Frequency terms	20
Matthews, Wright, & Yudowitch (1975)	Army enlisted personnel and officers	51	Adjective Phrases	141

Summary and Conclusions

Table VII-22 gives a summary of the studies conducted to show the perceived favorableness of word and phrases. As can be seen in the table, a large variety of subjects have been used in the studies. By looking at the tables presented in this chapter common words can be found across studies. Mosier (1941a) showed that the same word gets the same rating if it is repeated in a list, which implies that words have an inherent meaning. The fact that words have an inherent meaning or perceived favorableness independent of context and instrument used was supported by Dodd and Gerberick (1960) and Mittelstaedt (1971).

Chapter VIII

CONSIDERATIONS RELATED TO THE PHYSICAL CHARACTERISTICS OF QUESTIONNAIRES

This chapter considers four topics related to the physical characteristics of questionnaires: the location of the response alternatives relative to the question stem; questionnaire length; format considerations such as color, type size, spacing, and numbering; and the use of answer sheets.

Location of Response Alternatives Relative to Stem

Only two articles were found that pertained to the location of response alternatives relative to the question stem. Blumberg, DeSoto, and Keuthe (1966) had over 100 subjects rate well-known names on a variety of traits, using a nine point scale. They concluded that untrained raters can make relatively error-free ratings without being influenced by whether or not the "good" end of a graphic rating scale was at the left, right, top, or bottom.

The purpose of a study by Madden and Bourdon (1963) was to determine whether mean job evaluation ratings would differ as a function of seven variations in rating scale format. One of the variations included printing responses vertically or horizontally. Sixty basic airmen rated 15 occupations on nine job requirements for each format. It was concluded that the rating scale format was a determiner of the judgment of the raters in the sample .

Questionnaire Length

This section considers the effects of overall questionnaire instrument length on response rate, response inconsistency, and validity. Disagreement was found on the effect of length on the response rate of mailed questionnaires. Sletto (1940), in a 300 subject pretest of 10, 25, and 35 page mailed questionnaires, found no significant effect of length on response rate. Champion and Sear (1969) used 3, 6, and 9 page versions of the same number of questions spaced so as to affect apparent length. Mailing the questionnaires to 802 subjects, their results contradicted Sletto's findings since they obtained a greater response rate with the longer questionnaire. However, the overall response rate was only 35%.

Three other investigations concluded that the response rate for mailed questionnaires is greater for shorter questionnaires. Leslie (1970) concluded (without reporting data, however) that one or two page questionnaires improve the response rate for mailed questionnaires. Ford

(1968) found a slightly increased (but nonsignificant) response rate in a 1,556 subject test of the use of a printed, folder-type questionnaire, as compared with a larger appearing mimeographed, stapled format. One versus two page mailed questionnaires were tested by Bauer and Meissner (1963). They found that, in going from the one page to the two page format, nonresponse increased from negligible to over 5%. They also found that: absolute correctness of responses dropped from 53.5% to 47%; and nonsense answers increased from 1.5% to 5%. Their report, however, gave insufficient information to allow the reader to check the conclusions.

The effect of instrument length (in terms of total number of items) and other characteristics on response inconsistency was studied by Ace and Davis (1972). Using 177 college sophomores, they found that response inconsistency was only somewhat influenced by length and format, but considerably influenced by the type of scoring.

There have been a number of studies on the effect of instrument length on validity, but since they were concerned with cognitive and achievement tests, they were outside the main scope of this review. For example, Brokaw (1951), using six tests administered to 223 Air Force basic airmen to class them for training in technical specialties, found composite validity against course grades was .56 for half-length tests, .57 for full-length tests. Battery reliability of the half-length tests was .90, compared with .95 for the full-length tests. Since the tests measured reasoning and knowledge of facts, the results may not be generalizable to questionnaires as defined for this review.

In another study, Appel (1959) compared true-false and forced choice questionnaires, each administered to about 400 college students. He concluded that for longer forms the forced choice method is likely to result in greater validity, while for shorter forms the true-false method is likely to prove superior.

In conclusion, disagreement was found on the effect of length on response rate to mailed questionnaires, little information was found on the effect of length on response consistency, and nothing was found relating length to the validity of questionnaires as defined for this review.

Questionnaire Format Considerations

Little specific information was found related to questionnaire format considerations such as type size, spacing, color, etc. Sletto (1940) had 47 students rate the esthetic appearance of 10 different questionnaire formats, and found that preferences were not highly individualistic nor erratic. Wolfe (1956) discussed the effects of layout appearance, arrangement of questions and responses, and instructions. He noted differences, but provided no empirical data. Finally, Lehman (1967) reported that varying the length of a rating scale line from three and one-half to seven inches appeared unimportant in similarity ratings.

The Use of Answer Sheets

Several articles were located regarding the use of answer sheets, although this topic was not stressed in the literature review. Dunlap (1940) tested serially numbered, repetitively numbered, articulated, and unarticulated answer sheets in all combinations, using 20 groups of fourth and eighth graders. The sizes of the groups ranged from 251 to 364. His major conclusions were:

> 1. Marking articulated, repetitively numbered separate answer sheets is equally as satisfactory as underlining the correct response.
>
> 2. There is evidence that repetitive numbering results in more errors than serial numbering.
>
> 3. The use of articulated, serially numbered answer sheets is entirely satisfactory when compared with the results in using the underlining method.
>
> 4. The use of unarticulated but serially numbered answer sheets also seems justified. There was, however, a slight difference in results favoring articulated, serially numbered answer sheets.
>
> 5. Unarticulated, repetitively numbered answer sheets are somewhat less satisfactory substitutes for the underlining type of test than serially numbered, articulated sheets.
>
> 6. There is no evidence that the separate answer sheet cannot be used with children in grade levels at least as low as the fourth.
>
> 7. There is no evidence to support the contention that in a multiple choice test there is a psychological advantage in having the response indicated as close in time and space as possible (i.e., by underlining) to the decision as to the correct answer.
>
> 8. In summary, other things being equal, the use of articulated, serially numbered answer sheets is recommended, particularly if the test is short enough to enable all answers to be recorded on a single side of the sheet.

In a similar study, Faerber (1951) tested 230 students, finding a multiple choice test with a separate answer sheet more difficult than open answer, right/wrong, or multiple choice without a separate answer sheet when the tests were timed. When the effects of time were removed, the machine scored forms (all but the open answer) were more difficult than the open answer form. A different set of abilities for answering machine scored tests was hypothesized.

Bell, Hobb, and Hoyt (1964) compared a standard two page "fill in the mark" machine scored answer sheet with a new condensed one page answer sheet. For 1,048 civilian employees, the condensed sheet produced significantly lower scores, leading the authors to attribute the difference to the decreased type size. They concluded, however, that measures can be taken to compensate for the change. In a related experiment with 482 subjects using cross-out instead of fill-in answering on the condensed sheet, no significant differences were found between the one and two page answer sheets. The authors did not examine difference in subject familiarity with the two forms.

A comparison of answer sheets was also made by Dizney, Merrifield, and Davis (1966). Using an arithmetic test, they found that,in response to each of three questions,proportionally more students using the IBM 1230 format reported difficulty in using the answer sheet than did those students using the older IBM 805 format, although the answer sheets were similar. However, a statistical test of the scores of those reporting difficulty using the two formats indicated no significant differences.

In order to investigate the age range over which separate answer sheets could be used, Solomon (1971) tested 116 inner city fourth graders with three different answer formats for a reading test: answers within the booklet; separate hand scorable answer sheets; and separate machine scorable answer sheets. No statistically significant differences were found. For an older age group, Hart, Faust, Rowland, and Lucier (1964) recommended the use of optical scan and reusable booklets with graduated pages whenever possible. Their report, using a sample size of 2,160, was on a study of the attitudes of troops in the tropics.

In a study of problems related to the use of answers sheets, Swordes (1952) found that respondents frequently erred in using the last space on a multiple choice answer form when there were more spaces than actual choices. Precautions should, therefore, be taken, such as using the same number of distractors.

Although the studies reported above had to do with the use of answer sheets with achievement tests, the results would appear generalizable to the construction of questionnaires.

Chapter IX

CONSIDERATIONS RELATED TO THE ADMINISTRATION OF QUESTIONNAIRES

Considerations related to the administration of questionnaires are considered in this chapter since such matters are obviously of concern when questionnaires are constructed. The effects of instructions upon questionnaire results are first discussed, followed by sections on the effects of: various motivational factors; anonymity; administration time; characteristics of questionnaire administrators; administration conditions; and other factors such as bias and halo.

Effects of Instructions

Several studies discussed the amount of variance in responses due to variations in giving instructions. Some of the variance in instructions is unintentional, which was indicated in a study conducted by Belson (undated, a).In that study 236 tape recorded interviews were conducted, in which respondents were asked to use the semantic differential scaling system. The interviewers were told to deliver the printed instructions word for word. Analysis of the tape transcripts showed only 2% of the instructions were delivered word for word. Deviations from the instructions took the following forms: total phrases were eliminated with considerable ad libbing; and key words intended to focus the respondent's attention on some specific part of the instructions were frequently omitted or changed. The deliveries were rated for accuracy in presenting the 34 basic ideas in the instructions in the average delivery. As a result 28% of the key ideas were lost, mainly through omission. The variability of the interviewer performance varied substantially both across interviewers and within individuals.

Madow (1965) stated that the interviewer's attitude toward the question communicates itself sufficiently to the respondent to alter the meaning of the question. He concluded that the nature of the survey and the survey organization are determining factors in whether or not the interviewer must follow the interviewer schedule verbatim or may vary the wording.

Instructions are often varied in experiments to induce response sets. These experiments usually use standard instructions and instructions to fake. In a study by Winters and Bartlett (1966), a forced choice scale was constructed to provide independent measures of two types of response tendency, acquiescence and social desirability. The scale was administered under standard and faking instructions. Factor analysis yielded a social desirability factor under each instructional set, and an acquiescence factor only under standard instructions. Social desirability scores were observed to be orthogonal between instructional conditions. In another study conducted by Bartlett and Doorly (1967) using a forced choice scale measuring social desirability, the authors found that different instructional sets do affect the tendency to answer in a socially desirable way. Lederman (1971) administered two formats of the Thorndike Dimensions of Temperament to college students under regular directions and under instructions to give socially desirable responses. He found that the forced choice format produced low scale intercorrelations under regular directions, but under the desirability directions a common factor appeared. He also found that the same factor appeared in the questionnaire format under both regular and desirability directions. French (1958) found that significant differences were observed in most of the scales of the Edwards Personal Preference Schedule under different instructions.

Rambo (1968) found, using the Hinckley Scale of Attitude toward Negroes and a Likert scale, that the magnitude of the linear association between the scales was influenced by the instructions presented to the subjects. Frederiksen and Messick (1958) found that instructions altered mean criticalness set scores in the expected direction to an extent that it was significant on one of the three tests used, nearly significant on another, and nonsignificant on the third. And Bloxom (1968) found that mildly anger arousing printed instructions when compared with non-anger arousing instructions elicited more responses of negative self-report.

Jarrett and Sherriffs (1956) concluded that telling people to answer every item on a questionnaire or to omit an item if there is clearly no difference does not yield different results. Miron (1961) found that telling subjects to answer a retest of a questionnaire the same way they had answered the original questionnaire was superior to nonrecall conditions with respect to mean absolute test-retest deviations.

Berger and Sullivan (1970) examined an hypothesis that instructions emphasizing a respondent's importance in an attitude survey would result in a reduced number of "don't know" responses to the items. A 20 item questionnaire was administered to 180 undergraduates under three contexts: face-to-face interviews; telephone interviews; and group administration. Contrary to the hypothesis, the telephone interviews and group administration contexts yielded significantly more "don't knows" under the instructions emphasizing the respondent's importance than under the control instructions. There was no difference between instructional sets in the face-to-face context.

From the above discussion it appears that instructions do effect the responses collected by questionnaires. It also appears that more systematic research is needed to determine the range of variations in instructions that may affect the results given on questionnaires, and the effects of variations in respondent understanding of instructions.

Effects of Various Motivational Factors

In this section, the effects of various motivational factors are considered. The effects of a lack of respondent motivation will first be briefly considered. Attention will then be given to factors that affect the rate of return of questionnaires. Respondent preferences for certain item formats will next be reviewed, followed by a discussion of the effects of the behavior of the administrator on questionnaire response.

Effects of lack of respondent motivation. Some of the studies dealing with motivation discuss the effect of lack of motivation on responses to items. Flanagan (1955) tested the effects of motivation in a study of two groups. One group was reported as having high motivation: a group of Air Force Aviation cadets taking the Aircrew Job Elements Aptitude Tests. The other group was reported as having low motivation and consisted of seniors in their last two weeks of school or students involved in five days of testing. The incidence of patterned responses or random responses were higher in those groups where the motivation was suspect.

Hart, Faust, Rowland, and Lucier (1964) found that respondents who made three or more consistent errors in a study to assess the attitudes of troops in the tropics were more negative in attitude. Levy-Leboyer (1955) similarly found that omissions on achievement and intelligence tests were more due to motivation at the moment than to any persisting psychological trait. Finally Kendall (1954), in a study of factors which seemed to contribute to unstable responses people make to attitude questionnaires, found that shifts in the mood of the respondents and the degree of interest in or concern with the questions posed, affected results.

Factors affecting the rate of return of questionnaires. Some of the factors that affect the rate of return of questionnaires are reviewed below. Included are: the effect of ego involving the subject in the study; the use of advance letters, cover letters, and other techniques to stimulate the return of mailed questionnaires; and other factors.

Three studies examined the effect of ego involving the subject in the study. Slocum, Empey and Swanson (1956) found that efforts to establish an image of the social utility of a survey and to emphasize the special role of each respondent maximized the responses to a questionnaire and structured interview. Sletto (1940) found that three different cover letters did not significantly affect the rate of return of mailed questionnaires. One cover letter requested help on an altruistic basis, another on a challenge basis, and a third on a "help us" basis. In contrast to the Sletto (1940) study, Champion and Sear (1969) found that egoistic cover letters produced greater response rates than altruistic. This was found true especially in the case of lower class respondents. Cahalan (1951) conducted a study in which a questionnaire dealing with Army interests was mailed to 1,051 Army officers. The author felt the 84% return rate was due to the interaction of the institutional control applied by the Army, or the traditional responsibility of Army officers, with the wording of the cover letter which stressed responsibility and requested a return within five days.

In examinations of the effect of letters sent in advance of the questionnaire on response rates, Ford (1967) found that such letters significantly improved response rates. Myers and Hang (1967) also found that an initial letter increased response rates significantly, but Brunner and Carroll (1969) found that the initial letter did not significantly reduce interview refusal rate.

Glickman (1962) concluded that repeated administration did not have an adverse effect on the proportion of subjects returning questionnaires. Durant and Maas (1956) also found that people previously approached responded more readily a second time.

Several studies have been conducted to determine what techniques are useful to stimulate the return of mailed questionnaires. One technique involved including a stamped and addressed return envelope with the questionnaire. Ferris (1951) determined that including the stamped and addressed return envelope increased the response rate 53%. Clausen and Ford (1947), however, concluded that including prepaid return envelopes did not influence the return of questionnaires.

Another technique that was studied was the followup with reminders to complete questionnaires. Myers and Hang (1967) got a response rate of 28% for a group that had a followup letter mailed one week after the mailing of a questionnaire, and a 28.9% response rate from a group treated similarly but with no followup. Clausen and Ford (1947) also concluded that followup had no effect on response rate. Ferris (1951) similarly concluded that prodding with reminding postcards does not increase response rate. But Watson (1965) did find that followup postcards increased response rates from 30% to 40%, and two day followup mailing to the entire sample raised the response rate from 30% to 46%. Leslie (1970) recommended, without data to support it, the use of second, third, and fourth mailings and followup with personal calls to improve response rates.

A third technique which was studied involved the type of postage used to send the questionnaire. Watson (1965) obtained the same response rate using air mail and third class mailing. Clausen and Ford (1947) concluded that special delivery did not influence response rate. Champion and Sear (1969) came to the same conclusion.

The effect of incentives on return rates of mailed questionnaires has been studied by Brennan (1958). He tested the hypothesis that trading stamps would be an effective incentive to improve the response return rates in mail surveys. A questionnaire was sent out without incentives, with 50 trading stamps included, or with the promise of 100 trading stamps or 25¢ upon return of the completed questionnaire. The results showed no significant differences: the average return rate under all three conditions was approximately 27%. Watson (1965), however, received a return rate of 40% for a 10¢ incentive, of 41% when a packet of stamps was used, and of 48% when a 25¢ incentive was used.

Ferris (1951) found that responses were most frequently mailed on Thursday and Friday, least frequently on Saturday and Sunday. In turn, Leslie (1970) suggested mailing questionnaires so that they arrive on Thursday or Friday.

Simon (1967) investigated the effect of personally typed cover letters versus mimeographed form letters on response rate in mail surveys. He found that there was no overall clearcut advantage for personally typed cover letters in terms of response rate. He also noted the possibility that in some cases personally typed cover letters may reflect lack of anonymity and may therefore decrease response.

Ford (1968) demonstrated that printed, folder-type questionnaires generated higher responses than mimeographed, stapled questionnaires. This supports Leslie (1970) in the suggestion that long questionnaires be printed, not mimeographed, because printing reduces the length by two-thirds. However, Durant and Maas (1956) found that having to fill in two questions did not greatly increase cooperative response over having to fill in a 53 item questionnaire, which suggests that length may not be related to motivation for returning a questionnaire.

Respondent preferences for certain item formats. Several studies demonstrated subjects' preferences for specific types of item formats. In some cases this preference did not seem to have an effect on the results. Steinbock (1972) found that a format in which subjects rated themselves on given items on a nine point scale was more acceptable than a format in which they had to select items most or least like themselves. Zavala (1965) found that raters prefer forced choice formats using four favorable items from which they choose the items most characteristic of the person rated. Waters and Wherry (1961a) determined that subjects were more favorable toward a response format allowing them to indicate the degree of applicability of each statement in the forced choice pairs than they were towards other forced choice formats. Waters (1966) also reported that a subject's reaction to a forced choice scale is more favorable when some method is incorporated whereby he is given the opportunity to indicate the degree of applicability of each item to himself. Gaito (1962) speculated that forced sort Q-sorting techniques may adversely influence a subject's spontaneity. Turgut (1963) showed that 57% of a group's subjects preferred the format of Edwards Personal Preference Schedule, and 32% liked the Q-sort format. Jones (1968) showed that subjects clearly preferred multiple category options over two category options. Subjects also reported that multiple choice and true-false continuums were more interesting than the dichotomous true-false format. Hughes (1967) reported that a check list was preferred over the semantic differential when both were first administered to subjects, but preference for the semantic differential increased from 11% to 34% in a retest situation while the preference for the check list declined from 57% to 40%. Hughes attributed the increased preference for semantic differential to the respondents becoming more familiar with it. Matell (1970) suggested that in constructing Likert-type scales, that the number of steps should be chosen by the respondent's preference.

Effects of the behavior of the administrator on response. Several studies concluded that reinforcing behaviors of the interviewer, test administrator, or data collector have an influence on the responses collected. The effect of the experimenter's influence was studied by measuring the responses first graders from middle and working class families made under the conditions of reinforcing, neutral, and non-supporting atmospheres induced by the experimenter (Sgan, 1967). Results indicated that middle class children were more susceptible to the experimenter's influence and that, under a reinforcing atmosphere, they were significantly more apt to change their preferences. Stember and Hyman (1949) concluded that interested respondents were more subject to interviewer effects than uninterested ones. Wickes' (1956) findings suggested that such comments as "good" or "fine" and such actions as smiling and nodding by examiners have a decided effect upon test results.

Marquis, Marshall, and Oskamp (undated) reported that, although respondents liked the interview more when the interviewer was supportive, his manner had no effect on the accuracy or completeness of the responses. However, Marquis, Cannell, and Laurent (1972) reported that the interviewer's use of reinforcement increased the accuracy of reports from respondents who had not completed high school, and had the opposite effect on those who had. Field (1955) found that praised respondents in public opinion interviewing situations tended significantly to offer more answers than the unpraised ones. Praising respondents tended to reduce "don't know" answers, but praising did not increase insincere or dishonest responses.

In a study conducted by Hildum and Brown (1956), it was found that "good" proved to bias results in a phone attitude survey while"mm-humm" did not. Matarazzo, et al (1964) reported a 31% increase in subjects' average duration of single utterances when the interviewer said "mm-hmm" all the time the subject was talking. In a cross validation study there was an 84% increase in the mean duration of single units of interviewee speech. Dixon (1970) used subjects who were high or low on a social desirability scale in an experiment using reinforcement to increase self-referent statements. The reinforcement was done by the interviewer saying "good" after every sentence using "I" or "We." High social desirability subjects responded to reinforcement by increasing equally the frequency of both positive and negative self-referent statements. Low social desirability subjects did not condition, but continued to make more positive than negative self-references.

Effects of Anonymity

Several studies have been carried out to determine the effect of anonymity on questionnaire responses. Pearlin (1961) carried out a study in which he found that people selecting anonymity in filling out a questionnaire had different characteristics than those who did not. A questionnaire was administered to the nursing force of a large Federal mental hospital. Respondents were given the option of anonymity. It was found that those selecting anonymity were no more negative in their attitudes on a number of critical issues than were those who signed their questionnaires. Anonymous respondents were more subject to feelings of incompetence as reflected by their low scores on a measure of self-regard, by their reluctance to voice opinions at work, and by their reported difficulty in coping with the questionnaire. A second distinguishing characteristic of the anonymous respondents was their generally cautious view of people about them and their motives. Finally, it was shown that the anonymous respondents had less involvement and interest than signers in the issues covered by the questionnaire. Based on these findings, Pearlin suggested that anonymity in administration is useful, but for reasons other than the prevention of the arousal of fear or threat.

A few studies provide evidence that anonymity is affected by more than signing or not signing a questionnaire. Wiseman (1972) noted that questionnaires provide more anonymity than interviews. Metzner and Mann (1952) conducted a study in which a fixed alternative questionnaire was compared to an open-ended interview with 328 employees in an electric utility plant. The subjects were given an attitudinal questionnaire about their job with five scaled responses to choose from, followed two months later with a personal interview asking similar questions. The respondent's anonymity was assured both times. Blue collar workers were more confident of the anonymity of the questionnaires, while white collar workers felt the interviews were not less anonymous than the questionnaires. In general, the interviews yielded higher proportions of satisfied responses than the questionnaires.

Knudsen, Pope, and Irish (1967) collected data from samples of premaritally pregnant white women by three methods. The first sample anonymously completed questionnaires in their physician's office, the second sample was interviewed confidentially, and the third filled out questionnaires in the presence of an interviewer. Data suggested that in the interview situations the respondent was more likely to support the public and restrictive sexual norms that she assumed were adhered to by the interviewer. Lower socioeconomic respondents deferred to the norms represented by the higher status interviewers. In the private and anonymous questionnaire situation, the respondents more often answered to subcultural norms.

Simon (1967), in his article on personally typed cover letters versus mimeographed form cover letters in mail surveys, advanced the possibility that in some cases the personally typed cover letters may reflect a lack of anonymity to respondents even though it is assured because the letters are addressed to them personally.

Hamel and Reif (1952) explored the question of differences in response due to signing or not signing the Employee Attitude Questionnaire. They found that essentially the same responses were obtained for individuals in identified or anonymous groups. They speculated, however, that these results may have been influenced by the fact that the staff of a university organization administered the questionnaires and respondents were repeatedly assured that the questionnaire would only be used for confidential research purposes.

Dunnette and Heneman (1956) investigated the effects on attitude responses of the identity of the survey administrator. Two employee samples were selected randomly from the total work force of a large department store. The IRC Employee Attitude Questionnaire was administered to one group by an Industrial Relations Center staff member and to the second group by the personnel manager of the store. The group which was given the questionnaire by the manager responded more favorably to the attitude survey than the other group. The same group tended to give fewer and shorter responses to open-end questions than the employees who were given the questionnaire by the Industrial Relations Center staff member.

Klein, Maher, and Dunnington (1967) compared attitude survey responses between identified and nonidentified manufacturing employees made under two conditions of identification. One condition involved a face-to-face designation by the respondent's manager as to which group he was to be in (high threat), and the other involved a random allocation as the respondent entered the testing room (low threat). All subjects were assured confidentially of their responses and the nonidentified respondents were assured anonymity. A positive distortion in responses took place under both identified conditions, but significantly more under high threat.

Bartlett and Sharon (1969) determined the effects of several instructional rating conditions on leniency on a graphic and forced choice rating scale. Approximately 1,000 undergraduate psychology students rated their instructors under instructional conditions indicating that the ratings: will be anonymous and will be used for research purposes only; may be used for evaluation purposes; will be identified by having the rater place his name on the rating form; or will have to be explained to the ratee by the rater. A significant leniency effect was found with the graphic ratings which were to be used for evaluation purposes and those that had to be justified to the ratee. It was concluded that the forced choice scale was quite resistant to leniency bias, however.

Some investigators found no differences in anonymous versus nonanonymous conditions. Edwards (1957a) found that assurance of anonymity did not eliminate or drastically change the nature of the relationship previously found between probability of endorsement and social desirability scale value where the assessments were not made anonymously. Ash and Abramson (1952) concluded that the verbally expressed attitudes of college students, as recorded on scales relating to ethnocentrism, political-economic conservatism, and anti-Negro prejudice, were not biased in either a more 'pro' or more 'anti' direction as a result of the requirement that they sign the scales, thus identifying themselves. Gerberich and Mason (1948) found, in the administration of a questionnaire on academic background, plans, and study habits to 2,876 students taking a biological science course, that there were no significant differences between signed and unsigned questionnaires. As mentioned above, Hamel and Reif (1952) also found no differences in signed versus unsigned questionnaires. But Corey (1937) found, in the administration of a questionnaire on cheating, that mean scores reflected a slight but statistically insignificant tendency for more sympathetic attitudes toward cheating to be expressed on anonymous papers.

Some studies have indicated that different results were obtained in anonymous versus nonanonymous situations. Fischer (1946) gave a psychological problem check list to 102 female psychology students, first with signatures required, then a week later without signatures required. The results indicated that the mean number of problems listed did not vary significantly under the two conditions, but the mean number of serious problems listed tended to be significantly greater when signatures were not required. In a study conducted by Olson (1936) a personality test to measure emotional instability was given to two comparable groups of college women, one group remaining anonymous, the other group signing their names. Subjects reported significantly more feelings and symptoms with neurotic implications under anonymous conditions than when required to sign their names.

The effects of anonymous versus nonanonymous data collections seems to be related to the content of the data. Wiseman (1972) conducted a public opinion poll on current social issues using three samples of Boston households controlled for socioeconomic factors. Data were collected from one sample by mailed questionnaires, from the second by telephone interview, and from the third by personal interview. For eight of the ten questions, the results under the three conditions were similar; but for two questions, one on contraception and one on abortion, the results differed significantly. On the anonymous questionnaire more people seemed to be in favor of such programs than in either the telephone or personal interview. Wiseman concluded that sensitive issues involving socially accepted or rejected answers will effect more response bias in interviews than in questionnaires.

In the Klein, Maher, and Dunnington (1967) study described above, items themselves produced variable distortion. Items dealing with salary and with ratings of top management produced consistent positive distortions under identified conditions, whereas items dealing with work pressure and the subject's manager produced little or no distortion. Dunnette and Heneman (1956) also found that the amount of response distortion depended upon the content of the items comprising the questionnaires.

Rosen (1960), as a result of a study with college freshmen completing signed or unsigned questionnaires on the effectiveness of a reading program, concluded that, when respondent identification is essential for correlational or followup purposes, the straight-forward approach is preferable to a number coding system. For sensitive issues or where there is expected distortion, it may be advisable to use an anonymous questionnaire. The other articles discussed above appear to support Rosen's recommendation.

In summary, it appears that anonymity depends not only on unsigned questionnaires but also on the conditions under which the questionnaires are administered. In addition, it appears anonymity only makes a difference when information on sensitive areas is collected.

Effects of Administration Time

Most of the studies conducted to evaluate the effects of administration time were done using achievement and performance tests which are beyond the scope of this study. The data were not pertinent to this review as they were based on the number of right answers or total individual scores. The related topic of questionnaire length is discussed in Chapter VIII.

Miron (1961) in a study using the semantic differential did vary directions in terms of how much time the subject was to take to respond to an item. One group was instructed to mark all items at a fairly rapid pace and to attempt to recall and duplicate their markings on the immediate retest. A second group was instructed to proceed at a slow rate throughout the testing and to recall and duplicate markings on the retest. A third group was instructed not to try to recall original testing judgments and to proceed rapidly on the retest. The fourth group was instructed not to recall but to proceed slowly. Test-retest correlations were computed for each of the groups and were all uniformly high. The standard errors of substitutions were found to range between .24 and .32 for groups one and three respectively, with an average absolute deviation range of .10 to .14.

It appears from the lack of studies in this section that further research is needed on the effects of administration time on subject's motivation and on the effects of setting time limits for completing questionnaires.

Effects of Characteristics of Questionnaire Administrators

This section reviews the effects that certain characteristics of questionnaire administrators have on the responses received from the people completing the questionnaire. Some of the studies dealing with interviewers appear to be generalizable to questionnaire administrators.

Sex of the interviewer. Colombotos, Elinson, and Loewenstein (1969) studied the effect of the interviewer's sex on interview responses. They found essentially no difference in the reporting of psychiatric symptoms to male and female interviewers in a community survey. However, they did speculate that differences in response patterns according to the interviewer's sex may depend on subject matter as well as on the composition of the respondent populations and other characteristics of the specific survey situation. They recommended that the rationales commonly presented for having either male or female interviewers be critically reexamined. Thumin (1962) found that the percent of people admitting to insomnia differed significantly according to the sex of the interviewer. Twenty-two percent of the subjects interviewed by male interviewers reported having insomnia, compared to 13% of the subjects interviewed by females. No interaction effects between sex of interviewer and sex of subject was found. In a study conducted by Boyd and Westfall (1965a),it was found that women had better ratings as interviewers than men.

Race of the administrator. Most of the studies conducted to determine if the race of the investigator had an effect on responses involved questionnaires concerned with race. In a study conducted by Summers and Hammonds (1966) a Negro attitude scale was administered by two investigators. In a portion of the groups both investigators were white. In the remainder of the groups there was a Negro and a white investigator. The results indicated that socially acceptable answers to the Negro attitude scale were reported with greater frequency when one of the investigators was Negro. However, the phenomenon was more pronounced among certain strata of respondents than among others, suggesting that the effects should be viewed as the result of interaction between investigator and respondent characteristics. Similarly, Sedlacek and Brooks (1972) measured the attitudes of whites toward blacks with the Situations Attitude Scale. Results indicated that there were no measurable effects attributable to the race of the person administering the Situational Attitude Scale.

Sattler (1970) reported on a comparative review of studies that considered the effects of the experimenter's/interviewer's race on physiological responses, task performance, intelligence testing, personality scores, attitudes and preferences, speech patterns, interviewing, and psychotherapeutic or counseling relationships. One finding reported was that respondents more often gave socially desirable answers to interviewers whose race was different from theirs, particularly if their social status was lower than that of the interviewer and the topic of the question was threatening.

Trent (1954) hoped to find an effect of the investigator's race on responses by asking black and white kindergarten children to select their mother from three photographs of models (one black, one brown, and one white) using either a white or black investigator. He found:

1. When testing white children, a white investigator produced more responses to white and brown mothers; but with a black investigator, the choices were for a black or white mother.

2. The results differed little with black youngsters except that with the black investigator they chose brown or black mothers more often.

3. The black children avoided making any decision 25% of the time with a white investigator but not at all with a black investigator, while there were no evasions by the white children in either condition.

The results of the study by Schuman and Converse (1971) indicated that the race of the interviewer affected only responses to questions about militancy and hostility toward whites, and not responses to nonracial questions or questions about discrimination. Black interviewers obtained more militant answers, particularly from lower socioeconomic status black respondents. White reports to white interviewers indicated a higher level of militancy for upper than lower income blacks, while black interviewers received a fairly even distribution. Schuman and Converse attributed this difference to a "white effect" rather than to a "black effect." One other question on favorite entertainers showed differences by race of interviewer, indicating that the interviewers' race can establish different frames of reference even in nonsensitive areas.

In a study by Babatz (1967) 120 Negro undergraduates were administered the Test Anxiety Questionnaire under eight experimental conditions. It was found that Negro subjects tested by a Negro examiner reported less anxiety than those tested by a white examiner.

Other characteristics of administrators. Ehrlich and Riesman (1961) investigated the tendency of teen-aged female students to give socially desirable answers to authority figures. A socially desirable answer was defined as showing stronger ties to parents or other adults than to peers.

Less socially desirable answers were more often given to younger interviewers and to more flexible and less authoritative interviewers (as judged by personality scores). However, for interviewers over 53, personality did not make a difference as people over 53 were seen as authority figures regardless of their personality characteristics. Respondents under 16 did not show as clear a differential by age of interviewer as did those between 16 and 18. Sattler (1970) reported that the greater the disparity between the status of the interviewer and that of the respondent, the greater the tendency for biased responses. And Siegman, Pope, and Blan (1969) found that more productive responses were elicited by high than low status interviewers.

Atkin and Chaffee (1972) tested the ingratiating effect that an interviewer may have over a subject. In one study residents were asked about their opinions about firefighters. Half of the subjects were told their interviewers were firemen while the other half believed they were only students. In another related study mothers were asked their opinions of violence on TV. Half were told their interviewers were on a Federal committee investigating TV violence, while the other half were told their interviewers were students. The results showed significant differences between the two groups in each study, which suggests a subject will try to answer favorably in the eyes of the interviewer, if the subject can determine some means of response bias.

Quinn (1967) examined the hypothesis that performance raters would tend to rate subordinates higher who were most like themselves, using military officers. Results indicated that there was no evidence that performance ratings were influenced by similar characteristics of rater and ratee. Johnson (1958) had company interviewers rate job applicants, and the applicants rate the interviewer. He concluded that personnel selection is largely a matter of harmony of personal characteristics between the interviewer and the interviewee.

Some studies dealt with the experience of the interviewer related to the responses received. Smith and Hyman (1950-51) found that interviewers with more than a year of experience made fewer errors in recording data than those with no interviewing experience. Schyberger (1967) reported that the results of a study showed nonsignificant differences between interview completion rates for experienced and inexperienced interviewers, and that the training and experience of the interviewer had no effect on the number of deviations they made from the instructions. Boyd and Westfall (1965a) reported that all interviewers improved with experience, and training improved interviewers with a high school education but had little effect upon interviewers with a college education.

More research needs to be done on the characteristics of the administrators of questionnaires. No studies were uncovered related specifically to military personnel. For example, the military rank of the person administering a questionnaire may have an effect, as might whether the administrator is in the military or not.

Effects of Administration Conditions

The effects of questionnaire administration conditions was studied by a number of authors. In a study conducted by Hinrichs and Gatewood (1967), male technical employees in a large national organization rated their degree of satisfaction or dissatisfaction with various aspects of their work. It was found that conditions under which the survey was administered did have an effect on response. When employees were surveyed on their job location under the supervision of a company representative, there was a tendency to respond more favorable to a significant number of general opinion questions, particularly questions dealing with the company in general, than when they were permitted to respond to a questionnaire mailed to their home.

Green (1951) determined that a larger percentage of people attempted to fake on the Kuder Preference Record, the Guilford Inventory of Factors, and the Guilford-Martin Inventory of Factors when these tests were used for selection purposes than when they were administered to a control group. Rainio (1956) also studied the effect of the selection situation on responses to questionnaires. His experiment showed a significant difference between research and selection situations for various traits. In the selection situation there was a trend to higher scores on those variables shown to have higher correlations with the criterion. Heron (1956) designed an experiment to measure the effect of differences in test conditions. Applicants for the job of omnibus conductor were given a two part personality test covering emotional maladjustment and sociability. In one case it was administered as part of the application process along with a health examination. In another situation the test was administered to individuals after they had been hired. There was a statistically significant difference between the mean scores and variances. The group administered the questionnaire as part of the application process had a higher mean score and greater variance.

Several of the studies on the effects of questionnaire administration conditions were concerned with raters and their performance. Bayroff, Haggerty, and Rundquist (1954) studied the validity of rating related to rating techniques. Officer students served as a rater-ratee population using two types of graphic rating scales and two modifications of the forced choice technique. Results indicated that ratings earlier in a series were more valid than those at the end.

Freeberg (1969) studied the relevance of rater-ratee acquaintance in the validity and reliability of ratings. Unacquainted subjects worked in three-man groups under relevant and irrelevant acquaintance conditions. The subjects rated one another on scales that defined several cognitive skills. They were also rated on these same scales by observers who were dependent on visual information only and were unacquainted with the group members or the nature of the task being performed. Group members under the relevant acquaintance condition achieved consistently good validity ratings for all three cognitive areas, with the best validity rating on mathematical ability. Validity under the irrelevant acquaintance condition was nil on all scales. Observers achieved significant validity (although at lower levels than participating group members) only for ratings under the relevant acquaintance condition.

Shen (1925) found that when 28 subjects ranked each other on friendship and eight other traits (intellectual quickness, intellectual profoundness, memory, impulsiveness, adaptability, persistence, leadership, and scholarship) there was a tendency to overestimate friends on all traits except impulsiveness and to underestimate those rated as less intimate. Mayo (1956) concluded that for peer rating there is a substantial halo effect. This conclusion was based on a study of peer ratings of intelligence and effort with objective measures of both.

It is apparent that additional research is needed on the effect of administration conditions. Such research should include the study of fatigue factors.

Effects of Other Factors Related to Questionnaire Administration

Many things affect the data collected by questionnaires, raters, interviews, and observers. This section discusses those uncovered by the review of the literature: investigator bias; observer bias, halo effects, and the biasing effect of interviews.

Investigator bias. One of the main sources of bias comes from the researcher himself. Kornhauser (1947) discussed this problem of bias in research. He identified several biases: choice of subject matter; study design and procedure; unfair or loaded phrasing of questions; and interpretation and reporting of results. He felt the source of such biases are the researcher's relationship with the client, the researcher's personal involvement in a particular theoretical position or research technique, and those personal traits attributable to class, race, and political ideology. To reduce the impact of bias he felt that researchers need to be aware of such problems, need to seek critiques from independent sources, pursue public scrutiny through publication of reports, and continue to pursue technical improvement in opinion research.

Many things the researcher really is not aware of have an influence on results. Jensen and Schmitt (1970) designed a study to determine the extent to which responses to test items of the type frequently found in personality inventories would be influenced by the title associated with the test. An instrument was constructed and administered to eight treatment groups. Each administration differed primarily in the title the test bore. The dependent variables were measures of the tendency to lie, respond defensively, answer carefully, and complete questions. Subjects tended to lie and respond more defensively to titled tests than to a test having no title and administered under nonthreatening conditions. All other comparisons were not statistically significant.

Dillehay and Jernigan (1970) tested the hypothesis that biased questionnaires are effective in inducing changes in the subsequent opinions of respondents. Systematically biased and control questionnaires were constructed in a manner designed to elicit either harsh, lenient, or neutral opinions of respondents concerning the treatment of criminals. After answering one form of these treatment questionnaires, respondents registered their opinions on standardized attitude scales. The results indicated that the treatment questionnaires were successful in manipulating responses to lenient bias. Subjects displayed more lenient attitudes after exposure to the lenient form than after exposure to either the neutral or harsh forms of the questionnaire.

Question bias. Suchman and Guttman (1947) gave four suggestions for eliminating question "bias:" asking many questions on the same topic; determining by scale analysis whether questions ask the respondents about the same dimensions of opinion; asking "How strongly do you feel about this?" after each opinion question; and relating the content of opinion to intensity of feeling.

Observer bias. O'Leary (1973) and Skindrud (1972) studied observer bias in field studies. O'Leary found that simply informing observers of experimental hypotheses did not produce observational data consonant with those hypotheses. However, questionnaire responses following an experiment with different induced expectations did produce global data consonant with the experimental hypotheses. He also found, if observers are informed of the experimental hypotheses and the investigator provides daily feedback to them indicating how well their data support his hypotheses, observers will report data consonant with those hypotheses. Skindrud (1972) led three groups of observers to expect different outcomes from their observations. Even though the groups expected different outcomes, they were totally unbiased in their reports of deviant behavior in group comparisons. Failure to obtain evidence for observer bias in spite of the demonstrated manipulation of observer expectations was attributed to the precautions taken to assure high levels of observer accuracy.

Halo effects. Several studies discussed the halo effect, which is the tendency for trait ratings to reflect in part the rater's general impression of the person he is rating. Bingham (1939) reviewed the results from two examining boards responsible for rating 29 candidates for executive director positions in two Pennsylvania counties. He found the correlations between rating for the general category "Personal Fitness" and the ratings for specific traits such as voice, poise, freedom from bias, and ability to plan and organize to be positive and rather high. Johnson and Vidulich (1956) found that halo effect is a judgmental error rather than the effect of an objective correlation of traits. In their study one group rated five individuals, one individual per day on five traits, while another group rated five individuals on one trait per day. Johnson (1963) reanalyzed the data and found that the usual interaction between raters and individuals was found to be significant under both experimental conditions. Hence he concluded that the evidence for halo effect due to judging operations remains questionable. Bucklow (1960) concluded that, if items are constructed so "as to relate to clearly observable aspects of behavior which do not overlap," rating will be improved, although "halo" cannot be eliminated.

In a study conducted by Gordon (1972), the results indicated that neither the differential accuracy phenomenon (the situation where correct behavior is identified more accurately than incorrect behavior) nor the overall accuracy of ratings were related to the favorability of the rater's general impression of the ratee. He concluded that these findings make suspect the current practice of operationalizing leniency error by use of the average level of favorability of global rating. Bayroff, Haggerty, and Rundquist (1954) found that the average of a number of ratings was more valid than a single rating per ratee. Rappard (1950) also found that mutual arrangement between a number of raters is felt to enhance greatly the correctness of the rating.

Zavalloni and Cook (1965) concluded that ratings of unfavorable as well as neutral items are influenced by raters' attitudes. Extreme judges make fine discriminations at their own end of the scale and lump together the items at the other end. Falk and Bayroff (1954) concluded that the rater is the principal source of contamination in studies using ratings.

Biasing effect of interviews. Many studies have been conducted to show the biasing effect of the interview. Since unstructured interviewing is not within the focus of this review, only a few of these studies are discussed below to indicate the scope of the bias and some of the recommendations for controlling it.

In a study by Stanton and Baker (1942), five professionally trained interviewers obtained significantly more correct recognitions of previously exposed geometric figures when they knew the identity of the correct figures than when they did not. In contrast, in a study by Lindzey (1951) graduate students with training in interviewing methods failed to obtain significantly more correct recognition of previously exposed geometric figures when they knew the identities of the correct figures than when they did not.

Hanson and Marks (1958) reported that the factors leading to significant effects of the interviewer upon results are: relatively high ambiguity in the concept or wording of the inquiry; the interviewer "resistance" to a given question; and additional questioning or probing. Ferber and Wales (1952) reported that interviewer bias could exist without being apparent in an analysis of overall sample distributions. The direction of bias did not appear to be uniform. Cahalan, Tamulonis, and Verner (1947) concluded that the least interviewer bias was found in questions that could be answered "Yes" or "No". Shapiro and Eberhart (1947) reported that interviewer bias on attitude questions resulted from differences in the interviewing method used, differences in the degree of success in eliciting factual information, and differences in classifying the respondent's answers.

Back, Hill, and Stycos (1955), by analyzing the data reported from interviews in a fertility program in Puerto Rico, found reproducibility differences which were attributed to the interviewer and not to a response set of the respondents to four Guttman scales. Two "traits" were found among the interviewers which were negatively correlated: conscientiously completing the questionnaire, and understanding the study. The resulting effect is either a quality interviewer or a quantity interviewer, which should be decided by the type of data needed. Smith and Hyman (1950-51) concluded that interviewer expectations had a more powerful effect on the results (recording errors) than did the interviewer's ideological preferences.

Chapter X

CHARACTERISTICS OF RESPONDENTS THAT INFLUENCE QUESTIONNAIRE RESULTS

This chapter discusses various types of response bias. Response bias refers to the tendency of subjects to respond to questions in a pattern or set regardless of the content of the question. One hundred thirty-seven studies were searched indicating that the subject is recognized as an important aspect in questionnaire technology. These studies are discussed in terms of: item format biases; social desirability response set; acquiescence response set; extreme response set; the effects of attitudes on responses; and the effects of demographic characteristics on responses.

Cronbach (1950) and Horn and Cattell (1965) examined the disturbing effect of response bias on test reliability and validity. Fricke (1957) asserted that response bias could explain repeated findings that well-adjusted, successful persons obtain more abnormal scores on the subtle scales of the MMPI than maladjusted, unsuccessful persons. Rorer (1965), however, concluded that if a bias to consistently respond in a particular way exists, it would be eliminated by rewording the questions in the opposite direction. His results indicated that response bias could be attributed simply to the content of the question, therefore, as an intervening variable, would have only minor influence.

Most of the research on response bias does substantiate its existence. Nunnally and Husek (1958) demonstrated response bias by substituting randomly chosen foreign words for meaningful components of test items and then measuring the predisposition of subjects to give particular answers to these ambiguous questions. In a similar study McCord (1951) designed questions that could not be answered factually or truthfully by saying "yes," yet he found between 8% and 53% affirmative responses. Berg and Rapaport (1954) eliminated the questions altogether and had their subjects answer imagined questions; they found a great tendency among their respondents to choose culturally valued expressions such as "yes," "true," and"agree." Other researchers such as Webster (1960) have found high correlations between response patterns on personality inventories and personality measures like social alienation and schizoid functioning. Sudman and Bradburn (1974) have examined many possible sources for response bias and the effect this variable has on error in research. The remainder of this chapter will describe the studies done during the last twenty-five years in identifying the possible sources of response bias.

Item Format Biases

It has been shown that response bias is related to the format of the question and the methods of response available. Cronbach (1946) and Miklich (1966) have demonstrated how item ambiguity produces a recognizable pattern of response. Tajfel (1959) found that even abstract stimuli influence the physical attributes of items to be judged when no other information is available. Zajonc and Nienwenhyse (1964), however, found that the frequency of common words as answers played only a negligible role in response bias. Sax and Carr (1962), working with omnibus and subdivided test formats, and O'Dell (1962), studying the sequence of items, concluded that these format variables were significantly correlated with response bias.

Aschal (1958) concluded that response bias can be expected from all close-ended questionnaires where answers must be selected from two or more fixed choices. Jackson and Minton (1963) found that the forced choice format of selecting one of a pair of choices could eliminate a massive response set where some respondents tend to check many items from a list and others only a few. Considering true-false, multiple choice and card sorting methods, Van Der Veen, Howard, and Austria (1970) concluded that all three formats were relatively free of response bias; but, like Cataldo, Johnson, and Kellstedt (1970), they found card sorting to show the least effects of format response bias. One additional response bias, that of random markings, was noted by Flanagan (1955), and seemed to be related to the motivation of the subjects when they had no reason to take the test.

Social Desirability Response Set

Evidence has been found that a response set or style exists according to how favorably society would view the response. This type of social desirability response was found by Rugg and Cantril (1942) to be so powerful that subjects would not tend to deviate from social norms in their answers even though their behavior denied the opinion. Warren (1972) successfully trained some subjects to a particular response set but found that highly socially desirable items prevented facilitation. In an attempt to further define this factor of response bias, Fehrer and Strupp (1949) determined that prestige value had no effect on responses to job title preferences, and Krug and Northrup (1959) noted that on self-description inventories response time decreased as social acceptability increased.

The influence of social desirability was noted by French (1958) in scaling instructions that included the phrase "the Air Force way." When a respondent's job (Green, 1951) or his incarceration (Dubeck et al,)1971) depended upon his answers, there was a great tendency toward socially desirable responses. Wiseman (1972) found that anonymous questionnaires as opposed to personal interviews were necessary in order to surmount the social desirability response set on such socially sensitive issues as abortion or contraceptives. Heilbrun (1958) noted that under defensive conditions, subjects avoided unfavorable self-descriptive adjectives but did not necessarily increase selection of favorable adjectives.

Several authors (Edwards & Diers, 1963; Dixon, 1970; Potter & Tinkleman, 1970; Eysenck & Eysenck, 1963; Brod, Kernoff & Terwilliger 1964) have identified subjects with a high social desirability response rate. They found these respondents to give more true responses to neutral items, to be more susceptible to manipulation by social pressure, to more likely be introverts, and to score higher on a "lie" scale. Buss (1959) found that this response set was elevated with some subjects when given response choices styled like "trouble controlling," "must admit," and "tempted."

Faking or responding with socially desirable answers which are not true is a response error. Izard and Rosenberg (1958) gave instructions to their treatment group to try to fake their answers but found no significant differences between those and the control group's in a forced choice test. Several other authors (Leftwich & Remmers, 1962; Eisenberg, 1965; Bartlett & Doorley, 1967), however, obtained significant results showing fakability present in forced choice tests under varying instructional sets. Jones (1959) tried to neutralize faking by instructing subjects to do so and then establishing correlations of reliability with other tests. However, he was unable to achieve high correlations. Cliff (1968a) determined that faking responses as well as candid ones were simple functions of meaning space due to the great unanimity among the subjects concerning how to fake. Edwards (1957a) noted that even anonymity failed to eliminate the social desirability response set.

The forced choice instrument format has been studied for its susceptibility to social desirability response style. Silverman (1957) and Karr (1959a) found the forced choice method to minimize the effect of social desirability, while Krug (1958), Howe (1960), and Bernhardson and Fisher (1971) found the factor needing control in forced choice tests. Isard (1956) concluded that in forced choice formats, ambiguous items tended to be freer of social desirability response set than positively or negatively worded items. Due to the freer response choice of card sorting, Edwards and Horst (1953) and Edwards (1955) examined the method for social desirability effect but found it, too, needed controls to eliminate the bias. Hillmer (1958) found this response set to operate whenever the subject had the opportunity to respond in terms of it.

Krieger (1964) and Smith (1967) have both developed procedures for controlling or balancing social desirability by using loaded items in the test and then adjusting the subject's score.

Acquiescence Response Set

The response set to consistently agree, to say "yes," or to say "true," is called acquiescence. Upon comparing subjects taking attitude measuring tests, Lorge (1937) found correlations among those who marked "yes," "like," and "1" or "2" as well as correlations for those marking "no," dislike," and "7" or "8." Shipley, Norris and Roberts (1946) noted that judgement time was decreased when subjects were to choose the most pleasant color from many pleasant colors, or the most unpleasant one from many unpleasant ones. He concluded that this was an indication of an acquiescent response set. Other authors (Jackson & Messick, 1957; Mahler, 1962; Eysenck & Eysenck, 1963a; Foster, 1961; Quinn, 1963) have identified the acquiescence response set as a behavioral attitude to agree and accept even if subjects must alter their original opinions to do so. Elliot (1961) determined that acquiescence was highly dependent upon test construction and the respondent's aptitude, while Hanley (1965) found that acquiescence occurred with difficult rather than easy inventory material.

Gage and Chatterjee (1960) and Diers(1961) have further concluded that there is an opposing response set -- the naysayer -- which they found more valid than yeasayers. Wells (1963) identified both yeasayers and naysayers, but found more distortion in survey findings due to the former.

A still unsettled argument is whether or not acquiescence is simply a personality trait. Couch and Keniston (1960), Frederiksen and Messick (1958), Adams (1956), and Becker and Myers (1970) all pointed to the correlations of personality factors with acquiescence. In contrast, Foster and Grigg (1963), Eysenck (1962), and Findikyan (1969) found acquiescence unrelated to personality factors in personality measures, but conceded there may be a relationship in sociopolitical opinions or attitudes. To confound the matter, Cohn (1956) contended that the F scale is contaminated by an acquiescence response set, while Small and Campbell (1960) asserted that the relationship between conformity and the F scale is a function of content and not acquiescence.

Controls for acquiescence have been researched and some information is available on the response set's effect. Wells (1961) has detailed several design and statistical analysis procedures for eliminating the effect. Clancy and Garsen (1970) found that absolute scales of appeal were distorted by yea- and naysaying effects. Banta (1961) and Cloud and Vaughn (1970) concluded that item ambiguity increases an acquiescent tendency, but that, when it is minimized, balanced keying of items prevents contamination. Campbell, Siegman & Rees (1967) found that positive-negative reversal of items did not entirely eliminate the problem, but Findikyan (1969) concluded that reversal is an effective control if the items are not awkwardly worded. Falthzik and Jolson (1974) determined that a higher intensity of agreement is reached when items are positively stated than negatively stated. Knowles (1963), while finding the balancing of scales of dubious value to counteract acquiescence, did demonstrate that true-false questionnaires can be differentially prone to acquiescent response set.

There is a concern that social desirability and acquiescence are related in such a way that an individual with a tendency toward conformity will consistently reflect both biases. Several authors (Schultz, 1962; Stricker, 1962, 1963; Gloye, 1964; Liberty, 1965) have studied the relationship of the two effects and found no correlation. In two additional investigations, the two variables were studied but only one of them could be established to exist independently: Siller and Chipman (1963) found an acquiescent response set factor, and Winters and Bartlett (1966) found only an independent social desirability response set.

Extreme Response Set

Several studies have examined the possibility that an extreme response set exists where some individuals tend to consistently select exaggerated choices for positions. Rundquist (1950) found a low but significant correlation for preferred personality and interest items. Rather than use it as a predictive instrument, the author suggested attempts be made to eliminate the effect. Goldsamt (1972) found evidence that an extreme response style did exist on neutral content items, but that the effect was not generalizable due to its interaction with content-extreme-response-style. Mascaro (1968) examined the width of the response categories and extreme response, but found no significant correlation.

An unnecessary assumption is often made that a "don't know" or middle category response is lacking in demonstrative extremeness. Worthy (1969) and Ziller and Long (1965) presented evidence that this response is valid and can be related to dogmatism as a status-defense mechanism. Two studies (Adams, 1956; Cooper & Cowen, 1962) pointed out that extreme responses are not necessarily inhibited, and that a lack of inhibition would not explain the bias pattern.

Soueif (1958) found a positive correlation between extreme response style and intolerance to ambiguity. Levy-Leboyer (1955b),however, found that subjects who consistently omitted items were affected more by motivation than by a persisting psychological trait. Lucky and Grigg (1964) examined defensiveness and deviant responses, but concluded that outside self-description the two variables were unrelated.

Effects of Attitudes on Responses

A response bias attributed to an attitude is one which is influenced by the respondent's opinion, belief, or position. Shen (1925) recognized the disturbing influence that acquaintance had on raters and ratees. Hinckley(1932a),Prothro (1955), and Ferguson (1935a)found that by using Thurstone's methods of equal interval scaling, judges could rate items without being influenced by their own attitudes. Bruvold (1971), however, confirmed a competing hypothesis, that anti-attitude judges would rate unfavorable items higher and favorable items lower than pro-attitude judges. Other similarly disturbing attitude effects in response were found by Prothro (1957) concerning personal involvement of judges, and by Mogar (1960) involving high authoritarians in controversial social issues.

Explanations for these contrasting research conclusions came generally as controls developed. Kendall (1954) found unstable or changing responses were contributed to by shifts in the mood of the respondent, relative values among the possible choices, and the degree of interest present in the question. Kelley, Hovland, Schwartz, and Abelson (1955) found that blacks and whites in a competitive situation would make similar judgements concerning the social position of Negroes but when separated, blacks tended toward extreme responses. Zimbardo (1960) found no differences between pro- and anti-judges when well-structured sentences were used; but, as they became more ambiguous, the responses became more attitudinally biased. Upshaw (1962) noted that, if the judge's own position was outside the range of responses, bias would be evident. Rambo (1968) was able to shift judges' responses through attitudinal bias by instructing them to disregard their own opinions, which they could not do.

Other factors evidenced in attitudinal bias research are ego involvement (Paull, 1968), prestige value (Doncel, Alimena, and Birch, 1949), lack of attitude or position (Georgoff, Hersker, and Murdick, 1972), and task avoidance behavior (Weitman, 1964). In addition, Hill (1953) noted that inconsistent judgements decreased as psychological distances between items increased. French (1958) suggests that preference schedules be rescaled when attitudinal norms of groups differ greatly from the standard.

Effects of Demographic Characteristics on Responses

The final general area contributing to response bias is the effect of demographic characteristics. Such identifications as sex, age, race, or education have been examined to determine if similarities of such variables among respondents tend to be related to a response pattern. Roslow and Blankenship (1939) pointed out the theoretical need for designing questions with the respondent's background in mind. Johnson (1958) found that in personnel selection,the harmony of demographic characteristics played a major role in interviewer-interviewee relations. Schaie (1962) noted that factor matching in analysis depended on demographic knowledge of raters and responders. Jury (1971) determined that demographic characteristics reflected differences in workers' views of organizational-type variables, but that they were not related to individual-type variables.

Socioeconomic class has been identified by Soueif (1958) and by Clancy and Garsen (1970) as an influence of bias in response patterns. Sgan (1967) found middle class children to be more susceptible to experimenters' influence than lower class children. Race was found to be an identifying factor in extreme response rates by Sherif and Hovland (1953), while Sattler (1970) noted that response bias increased in interviews as racial disparity grew. Another characteristic bias, level of education, has been shown to relate to a decrease in acquiescent response style (Falthzik and Jolson, 1974), and to an increase in nonacceptance of causal explanations in ambiguous situations (Nunnally & Husek, 1958).

Several authors have identified other demographic characteristic variables such as age, religion, intelligence, sex, marital status, parenthood, nationality, urban or rural residence, income, rank, and experience. Such variables have been correlated with biases found in response consistency (Hart, Faust, Rowland & Lucier, 1964; Dakin & Tennant, 1968; Goldsamt, 1972; Flyer & Carp, 1962; and Sicinski, 1970). Aschal (1958) and Wells (1963) found correlations between acquiescence and demographic variables. Quinn (1967) in a study contrasting previous research, found no relationship between several demographic characteristics of raters and ratees and their ratings. Bauer (1947), Ferber (1966), and Ognibene (1973) found common characteristics of youth and less education among nonrespondents in mail surveys.

Other studies have explored more removed characteristics searching for significant differences in responses. Bayroff, Haggerty, and Rundquist (1954) examined "hard raters" and "easy raters" but found no differences in validity. High and low "feeling" persons were found by Frisbie and Sudman (1968) to make long answers on open-ended questionnaires. Ferber (1956) suggested that survey researchers determine the state of knowledge of their sample to avoid a response bias by persons ignorant of the issues and by persons misinformed about the issues. Two studies of military personnel (Hollis, 1954; Gilbert, 1956) indicated that the influence of occupational environment may be related to a bias against criticism and for acquiescence.

Summary and Conclusions

Response bias is an error factor in questionnaire technology due to a pattern of answers made by the respondent that appear to be related to extraneous variables. Several areas of origin for response bias have been studied which are grouped in this chapter into six categories.

1. Format biases are responses influenced by the question stem or response alternatives. Sequence and fixed choice responses have been related to this bias.

2. Social desirability has been well identified as a response set where persons answer according to the norms they believe society condones. The faking of responses on questionnaires contaminates the results, and controls must be designed to prevent its operation.

3. Acquiescence is the bias demonstrated by yeasayers who tend to respond more often agreeably than disagreeably. Some dispute remains over this bias as to whether it is actually a personality trait.

4. The extreme response set refers to the pattern of answers persons make which tend to be unevenly distributed toward one or both poles. As with acquiescence, some research indicated that this response style may also be a personality description.

5. Attitudes may influence responses in identifiable patterns. Opinions and beliefs seem to be related to a response bias.

6. Demographic characteristics have been shown to be related to response bias. Education, age, social class, etc., have been found influential in a response pattern especially noted by consistency.

Research during the last twenty-five years establishes a very strong case for the existence of response bias. Studies documenting its origins in social desirability, questionnaire format, and demographic characteristics are numerous. More evidence is needed to confirm that acquiescence, extreme response set, and attitudes are actually biases and not personality traits. None of the control measures examined thus far, including changing wording direction, balancing scales, using card sorts, forced choice, or open-ended designs, or loading questions, have convincingly eliminated response bias. More detailed identification and control methods are areas of needed further research in response bias.

Chapter XI

CONSIDERATIONS RELATED TO THE EVALUATION OF QUESTIONNAIRE RESULTS

Considerations related to the evaluation of questionnaire results was another area not stressed during the literature review. Some articles were reviewed, however, that pertain to the scoring of questionnaire results, the properties and uses of ipsative scores, and data analyses.

Scoring of Questionnaire Results

Practical considerations. Erdos noted (1948b) some points that are often forgotten until too late: that both time and money can be saved by planning the questionnaire in line with tabulation requirements. He used sample questions to illustrate the relationship between order of questions and tabulation, and how phrasing of questions, sequence, and layout can affect tabulation time. He also pointed out that whether data are to be tabulated by hand or by machine is an important decision and should be made in advance. The precoding of responses whenever possible was also recommended.

Quite early, Bass and Wurster (1956) described the use of IBM Mark Sense cards to put data on punched cards. They noted the procedure avoids the expense and difficulty of coding and keypunching large volumes of raw data. Of course, the use of Mark Sense cards has been largely replaced by one of a number of optical scanning procedures allowing the processing of regular sized answer sheets and booklets.

Lyman (1949) examined the assumption that items in a multi-scale inventory should be scrambled, even when the items are "obvious." He compared scrambled items and items blocked according to scale in a school attitude survey. Two high school senior classes were given the tests, one half of each taking the alternate version which was followed two weeks later by the other version. Test scores revealed no statistically significant differences, leading the author to conclude that blocking items may be preferable due to its greater ease of scoring.

Other considerations related to scoring questionnaires. Methods of scoring questionnaires, especially attitude scales, were discussed by a number of authors. For example, Kundu (1960) suggested a method for scoring responses on three point attitude scales. Assuming a non-normal distribution of attitude scores and non-neutral trends in attitudes, the "neutral" responses are broken into positive or negative and the responses are scored with the help of average group trends and weighted scores of the individual responses. Peabody reported (1962) that there is a justification for scoring items dichotomously according to the direction of response, as is done when bipolar scales are analyzed in terms of the proportion of responses in each direction of the basic dichotomy. The justification is based upon results he obtained that indicated composite scores reflect primarily the direction of responses, and only to a minor extent their extremeness. (He also noted that since extremeness scores are reliable and are largely unreflected in the usual composite score, they may have quite different correlates of their own.) Matell (1970), who investigated the psychometric characteristics of Likert-type rating scales consisting of two through 19 steps, found that, by collapsing the steps into two or three measurement categories for analysis, no lack of precision resulted. Schuessler (1952), however, raised doubt about the validity of combining response categories in successive approximations of scalability. He showed irregularities between analyses of a questionnaire form in which an "uncertain" response was permitted and combined as an approximation, and a second questionnaire form in which the "uncertain" response was not permitted. Odesky (1967), working with paired comparisons with a no preference option, suggested the advisability of either dividing no preference responses proportionate to preference responses, or disregarding them altogether. The basis of this suggestion was that respondents who claim neutrality appear to exhibit the same preference patterns as those who express a preference.

Two methods of scoring rating scale data to approximate forced choice results were reported by Karr (1959a; 1959b). One was called the difference method, and was designed to have maximum stability. The other, called the zero-one method, was designed to match as closely as possible the method of scoring for the forced choice format personality inventory with which the scoring methods were being compared. Karr concluded that by using any one of several methods of scoring or transforming self-rating scale raw scores, it is possible to approximate dyadic forced choice results with considerable saving in administration time, and a small gain in test-retest reliability.

It was hypothesized by Schaie (1963) that the concurrent validity of questionnaires can be increased by the use of item weights obtained by expert scaling, instead of by using conventional unit weights. The results, using a high school personality quiz, showed only low magnitude increments in validity, however.

Several authors reported on the use of intensity scores as distinguished from content scores. Guttman (1947a) showed how intensity scores can be obtained by either the fold-over technique or the two-part technique. The fold-over technique involved weighting extreme responses (positive and negative) as 2, moderate responses as 1, and neutral responses as 0, and summing these for an intensity score. The two-part technique ascertains an intensity score simply by following each question with the query "How strongly do you feel about this?" which is also answered on a scale. Goldsamt (1972), working with content-free stimuli, however, concluded that dichotomous scoring methods are equivalent to intensity scoring methods.

A rating-scoring technique for evaluating free response answers was described and illustrated by Canter (1953). The technique involved using raters to sort responses to each of four questions into a seven category forced normal distribution. The score for each respondent was the sum of the category numbers assigned by the raters over the four questions. Interrater reliabilities of .85 and .88 were reported.

A procedure for correcting the influence of social desirability (SD) response set in opinion research was reported by Smith (1967). Several SD items, to which it is assumed the true response is known (e.g., "Do you like everyone you meet?") are included in the questionnaire. The SD score from these items is then correlated with each of the other items on the inventory. The responses on those items with a statistically significant correlation can then be corrected, by moving the response one or more steps from the socially desirable response, to give a more accurate result.

Properties and Uses of Ipsative Scores

During the literature review, some attention was paid (but not an exhaustive review) to the topic of ipsative scores since their properties are not well known but should be understood by those who design questionnaires. It was in 1944 that Cattell (1944) noted that psychological measurement could be expressed in three kinds of units: interactive, normative, and ipsative. Interactive units are exemplified by the typical "raw" units of psychology, where one is measuring the interaction of the individual and his environment. Interactive units are neither dependent upon any other scores of the individual measured nor upon the scores of other individuals. Normative units are interactive measurements relative to a group of persons, or in terms of a population of measurements provided by a population of persons. Hence, the score of any given individual is dependent upon the scores of others in the population. Most scores in behavioral measurement are of this type. Ipsative units, on the other hand, are interactive measurements in terms of a population of measurements within an individual. Hence, the score for an individual on a variable is dependent upon his scores on other variables.

Several derivations of the three forms given above are possible. Where interactive measures are first scored normatively and then scored ipsatively, "normative ipsative" units are produced. "Ipsative normative" scores are also obtainable when ipsative units are themselves treated normatively. As Cattell (1944) pointed out, however, ipsative normative scores are not the same as normative ipsative scores.

Ipsative scores can be obtained in one of two ways: arithmetically through the use of various scaling procedures; and experimentally. Experimentally ipsative scores are produced, for example, by the forced choice technique, the use of paired comparisons, and the Q-sort. Examples of well-known tests that produce ipsative scores are the Allport-Vernon Study of

Values, the Edwards Personal Preference Schedule, and the Kuder Preference Record. A set of attribute measures is defined as ipsative when the sum of scores over all attributes is a constant for each entity.

The properties of ipsative measures were investigated by Clemans in 1956 and reported in a Psychometric Monograph in 1965. Some of his major findings, based on arithmetically ipsatized scores, and related implications and recommendations are given in detail below since, as previously noted, the properties of ipsative scores are still not well known.

1. There is always a set of raw or absolute measures underlying an ipsative set. They may be very difficult or impossible to obtain, but in theory they are there.

2. There will be a large number of negative values in any ipsative correlation matrix.

3. Ipsative intercorrelation matrices are nonbasic or singular and thus have no regular inverse. Hence, if regression weights are to be determined for a complete set of ipsative variables, special procedures (such as iterative techniques) will have to be employed or one of the ipsative variables will have to be deleted.

4. The least-square estimate of a criterion using all the variables of an ipsative set is identical with the least-square solution with any single variable deleted, regardless of the validity coefficient of that variable.

5. Ipsative scores must always be interpreted as relative and not absolute measures. Ipsative variables are highly interdependent.

6. Except in one special case, the ipsative multiple correlation is always less than the multiple correlation for the same variables prior to ipsatizing. Whenever possible then, absolute measures should be used for measuring attributes of behavior.

7. If the underlying absolute measures have zero correlation with each other, or they all correlate to some constant degree, the ipsative intercorrelations will all be a negative constant value determined only by the number of variables. This again shows the high interdependence.

8. Under certain restrictions, the ipsative covariance matrix and the first centroid residual of the absolute measures are identical, and the property seems to hold very well even without the restrictions. This is the same as stating that a tremendous amount of information is missing in the ipsative set. The fact that this information is missing from the ipsative set will make it next to impossible to make anything psychologically meaningful out of a factor analysis of such data, other than a determination of the rank of the matrix.

9. If the means and variances of a set of scores were not equated prior to ipsatizing the resulting scores will have little meaning. Although the means can be adjusted after ipsatizing, the variances cannot. Hence, the ipsative test maker must be as certain as he can that equal variances are maintained for the absolute scales underlying an ipsative set even though the absolute measures cannot be observed. How this is to be accomplished is unknown.

10. The magnitude of ipsative scores must never be confused with the magnitude of the absolute measures for the same set of variables. It is quite possible that a person with a low ipsative score on a particular trait may actually possess more of the trait than a person having a much higher ipsative score.

11. Although nonipsative measures contain more information than ipsative measures, it is not usually an easy task to develop absolute measures that correspond to the variables in an ipsative set. It was the difficulty of obtaining valid absolute measures that lead to the development of some of the available ipsative instruments, such as attempts to eliminate the social desirability factor. Hence, some traits that may be relatively easily compared using ipsative techniques may be very difficult to assess validly using instruments designed to yield more direct or absolute measures.

Seven other studies were reviewed that also investigated the properties and uses of ipsative scores. Block (1957) reported the results of two studies in which ipsative ratings, treated normatively, were correlated with corresponding normative ratings in a test of the functional equivalence of the two forms of measurement. Both of the analyses showed an almost complete equivalence between the two methods. Some of the advantages of the ipsative approach were also presented.

Horst and Wright (1959) reported on the comparative reliability of an arithmetically ipsatized rating scale and its experimentally ipsative counterpart. The experimentally ipsative scores were obtained from the forced choice format Edwards Personal Preference Schedule, the individual items of which were also administered in rating scale form to obtain interactive scores. The interactive scores were standardized by variable to produce normative scores, and these were then arithmetically ipsatized over persons to produce normative ipsative measures. It was found that the average reliability of the variables for the arithmetically ipsatized rating scale form was .87, while for the experimentally ipsative scale it was .78, even though the administration time for the rating scale was only about one-third that of the EPPS. They concluded that any advantage which the forced choice type of self-appraisal instrument may have over the arithmetically ipsatized rating scale must be other than that of greater reliability. They also suggested that other possible advantages be investigated in further research. In a related report, Wright (1961) compared the three types of measures with respect to their intercorrelations and factor structures. The data suggested that the normative approach provided more significant measures since one less factor was required to extract all of the approximate reliable variance from the intercorrelations of the experimentally ipsative units than from the normative and arithmetically ipsative scores. An inspection of the unrotated factor patterns showed that the first normative units factor was not adequately matched by either a normative ipsative or an experimentally ipsative factor. It was tentatively identified as a factor of social desirability, the attempted minimization of which was the reason the EPPS was developed in forced choice format.

The equivalence of ipsative and normative personality measures was also studied by Heilbrun (1963) with regard to interscale correlation and relative validities. Using normative check lists and forced choice ipsative Q-sorts, the results were interpreted as supporting the use of ipsative measures for normative predictions. Concerning reliability, Tenopyr (1968) noted that the practice of resorting to stability coefficients as reliability estimates for ipsative scores is not a satisfactory method by itself since these coefficients are subject to scale interdependency. He suggested that the recommended practice for establishing the reliability of ipsative inventory scales should involve the establishment of internal consistency and stability for the scales prior to putting them into the forced choice form.

Two reports mentioned the "degree of ipsativity." Smith (undated, but around 1965) reviewed the relevant literature describing the mathematical and empirical properties of ipsative and nonipsative measures. The review led to the explication of a simple procedure for quantifying the "degree of ipsativity" in psychological measurement instruments. After evaluating several published research studies against the index, he concluded that purely ipsative test instruments possess such extensive psychometric and statistical limitations that utilization of such instruments is not advisable. Hicks (1970) came to the same conclusion in what could be a later report on the same study. He went on to suggest, however, that ipsative tests should be used only in situations where it has been demonstrated that: significant response bias exists; this bias reduces validity; and an ipsative format successfully reduces bias and increases validity to a greater extent than do nonipsative controls for bias. Since Hicks felt that little of the research utilizing ipsative measures fulfilled these requirements, he believed that it is necessary to reevaluate thoroughly the extensive body of research that has used purely ipsative forced choice tests and that have employed statistical techniques predicated upon assumptions which such instruments necessarily violate. It may be noted that the conclusions of both Smith and Hicks are somewhat more extreme than those of Clemans (1965).

Data Analyses

Generally, reports on data analyses were beyond the scope of this review unless specifically connected with some aspect of questionnaire construction.

Most that were located were discussed above in other chapters in conjunction with their related topic. Four articles, however,may be noted here.

Stevens (1946) pointed out four kinds of measurement scales: nominal, ordinal, interval, and ratio. Appropriate statistical analyses are associated with each. Hence, the data analysis limitations of various forms of questionnaires should be considered before an instrument is designed. For example, from a power of the statistic point of view less can be done with open-ended questions than with ranking questions.

A statistical measure most appropriately used in conjunction with the method of paired comparisons was reported by Balinsky, Blum, and Dutka (1951). Called the coefficient of agreement, it enables the experimenter to measure the degree and test the significance of agreement among observers as to their preferences for a series of items offered for consideration. It can be readily used in the construction and testing of attitude and opinion scales.

Litwak (1956) points out that ad hoc rules on question wording can be systematically defined by the constraints of latent structure analysis. And Reynolds (1966) attempts to determine the degree of difference between two ratings required for statistical significance with samples of varying sizes.

Chapter XII

RECOMMENDED AREAS FOR FURTHER RESEARCH

This chapter contains recommendations for further research based upon a lack of empirical research or contradictions in results of the studies reviewed in the previous chapters. The section headings used correspond to the previous chapters.

Advantages and Disadvantages of Various Types of Questionnaires

1. Because of the lack of stress in this review on mail questionnaires, only a few articles were discussed in Chapter II. Additional information could probably be found by extending the literature search.

2. More research appears to be needed on the benefits, validity, and reliability of combinations of questionnaire methods, for example interview and self-administered questionnaires.

Selection of Questionnaire Items to Be Used

1. More research appears to be needed on the comparison of ranking and rating techniques. For example, there is some evidence that conclusions based upon a single judge differ from those based upon multiple judges. Also, more studies need to be designed where the items to be ranked or rated are as comparable as possible.

2. Contradictary evidence was obtained regarding the comparison of ranking and paired comparison, which suggests further research.

3. More studies need to be conducted on the comparison of rating scales and forced choice items, where identical items are used in both forms.

4. Since few studies were located on the comparison of rating scales and card sorts, rating scales and semantic differential items, and rating scales and check lists, more studies can be carried out in these areas.

5. More research is needed on the comparison of multiple choice items with other item types.

6. A more critical and detailed review is needed regarding issues related to forced choice and paired comparison items.

7. A more extensive literature review regarding the use of the semantic differential might be in order.

8. Because of the focus of this study, few articles were located on card sorts and projective items, suggesting that a more complete literature review could be conducted.

9. Few studies concerning check lists, open-ended items, rearrangement items, and matching items were uncovered, suggesting another possible area of additional research.

Comparison of Scaling Techniques

1. This review did not stress scaling techniques on which many articles have been written. It is suggested that a review of the literature could be done stressing just scaling techniques in regard to questionnaire construction.

Effects of Variation in Presentation of Questionnaire Items

1. Even though the review of the literature revealed that pictures can be effectively employed in questionnaires, the review should be extended to determine other modes of item presentation that can be used in questionnaires (i.e. use of tape recorders or physical objects).

2. Follow-up research is warranted in the area of question stem wording. Many important issues have been raised by the studies presented, but there has been little systematic pursuit of the issues to a conclusion.

3. Since no research studies were uncovered which examined the wording of response alternatives, research needs to be done in this area.

4. More attention has been devoted to measures of item difficulty than to the effects of item difficulty on questionnaire responses. Additional attention needs to be focused on item difficulty response tendencies such as acquiescence, "don't know," and "no responses."

5. The effects of the length of a question stem is an under-researched area. Studies should be conducted to the point where generalized conclusions can be made.

6. Experiments controlled for subjects' characteristics, topical area, scale length and instrument should be conducted to determine the effects of the order of response alternatives.

7. No research was uncovered relative to adjective location in the stem of the question versus adjective location in the response alternative. Such research needs to be done.

Number of Response Alternatives and Response Anchoring

1. Additional research in the area of the optimal number of response alternatives to use is warranted. This research should cover: the different types of rating scales; various topical areas of research; and subjects with different ability, educational and sociodemographic characteristics. From such research information would be available regarding the optimal number of response alternatives to employ for any specific type of investigation situation.

2. Additional work needs to be done on the use of balanced versus unbalanced scales.

Order of Perceived Favorableness of Commonly Used Words and Phrases

1. Even though extensive work has been done on the order of perceived favorableness of commonly used words and phrases, individual investigators may want to determine the order of perceived favorableness of words which are not included in the lists in Chapter VII and which are commonly used.

Considerations Related to the Physical Characteristics of Questionnaires

1. Research needs to be conducted in regard to the location of response alternatives relative to the question stem.

2. Additional studies need to be carried out to determine the effect that the length of a questionnaire has on both the respondents' motivation and on the return of mailed questionnaires.

3. Another possible area of research would be to determine the relations of questionnaire length to response consistency and validity.

4. Systematic research needs to be done on the physical appearance of questionnaires including type size, spacing, color, type of paper and the use of pictures.

Considerations Related to the Administration of Questionnaires

1. More systematic research is needed to determine the range of variations in instructions that may affect the results obtained from questionnaires and on the effects of variations in respondent understanding of instructions.

2. Further research is needed on the effects of administration time on subject's motivation, and on the effects of setting time limits for completing questionnaires.

3. No studies were uncovered that were concerned with the effects of the administrators of questionnaires in the military setting. For example, the military rank of the person administering a questionnaire may have an effect, as might whether the administrator is in the military or not.

4. It is apparent that additional research is needed on the effects of administrative conditions. Such research should include the study of fatigue factors.

Characteristics of Respondents that Influence Questionnaire Results

1. Controls for all types of response bias need research.

2. The question of "Is attitudinal bias and characteristics bias automatically eliminated with stringent sampling controls or must each instrument take this into account?" needs to be resolved.

Considerations Related to the Evaluation of Questionnaire Results

1. A more extensive review should be made of work related to the properties and uses of ipsative scores and research should be undertaken to fill the gaps since procedures and techniques producing such scores are in wide use.

2. The literature review could be expanded to include scoring and data analysis, as related to questionnaire construction.

General Recommendations

1. The literature review could be expanded to cover citations that were not abstracted.

2. The present bibliography could be refined, maintained, and updated. Possibly it could be computerized so that requests for needed information could quickly be answered.

3. The present literature review could be reviewed by senior consultants in the field and expanded or modified on the basis of their suggestions.

4. Many conclusions presented in this review could be tested in relation to the military situation.

5. An attempt could be made to collect data about relevant issues on questionnaire construction from groups who routinely administer questionnaires but who might not publish their findings.

Bibliography

BIBLIOGRAPHY

Over 500 references are cited in this report. These references are asterisked in the following bibliography of over 2,000 citations. For those citations for which abstracts were prepared or were available, the information explained below is also presented:

1. The first code at the top center of each citation indicates the perceived relevance (H for High, M for moderate, N for negligible, NA for those found to be not applicable) of the citation, as estimated from the title (T), abstract (A), or report (R).

2. The second code at the top right of each citation indicates the technical categories which the citation appeared to address, from the following list:

1. Type of Instrument
2. Response Form
3a. Number of Response Alternatives
3b. Order of Response Alternatives
3c. Order of Questions
3d. Adjectives in Stem vs. in Response Alternatives
3e. Location of Response Alternatives Relative to Stems
3f. Response Anchoring
3g. Miscellaneous Format Considerations
4. Clarity
5. Instrument Length
6. Perceived Favorableness of Adjectives
7. Administration of Questionnaires
8. Evaluation of Questionnaires
9. Personnel Characteristics
10. Personnel Attitudes
11. Motivation
12. Bias
13. Other Sources of Error
14. Questionnaire Development Processes
15. Miscellany
16. Reference Source
17. Questionnaire Construction Guidelines
18. Found to be Not Applicable

3. Next the citation is given, in standard American Psychological Association format.

4. Following the citation itself are selected descriptive words or phrases further identifying the subject of the citation.

5. Finally, the source of the abstract is noted.

Abbott, R.D. Stylistic response set variance and trait inference from the study of values. Psychological Reports, 1970, 27, 911-914.

Response Bias 12

Psychological Reports, 27, p. 911 (Rev. from rept.) R-M

Abeles, H. F. A facet-factorial approach to the construction of rating scales to measure complex behavior. Journal of Educational Measurement, 1973, 10(2), 145-151.

Abelson, R. P. A technique and a model for multi-dimensional attitude scaling Public Opinion Quarterly, 1954, 18, 405-418.

Abrahams, N. M. & Lacey, L. A. The Navy Adjective List as a predictor of enlisted retention. U.A. Naval Personnel and Training Research Laboratory, 1972. Research Memorandum 73-2-1-7.

Personality Measures 18

Psych.Abst.., 49, #12129 R-N

* Abrams, J. An evaluation of alternative rating devices for consumer research. Journal of Marketing Research, 1966, 3, 189-193.

Instrument Format, Multiple Choice Items, Rating Scales, Response Alternatives, Validity 3f

ORA R-H

Abul-Ela, A. -L., A., Greenberg, B.G. & Horvitz, D. G. A multi-proportions randomized response model. Journal of the American Statistical Association, 1967, 62, 990-1008.

* Ace, M.E., & Dawis, R.V. The contributions of questionnaire length, format, and stype of response inconsistency. Educational and Psychological Measurement, 1972, 32(4), 103-111.

Instrument Length, Instrument Format, Investigator Error, Scoring	3a, 5, 8
ORA	R-M

Ace, M. C., & Dawis, R.V. Item structure as a determinant of item difficulty in verbal analogies. Educational & Psychological Measurement, 1973, 33, 143-149.

Achievement Measures	18
ERIC Document Reproduction Service ED 073 050	A-N

Adams, C. R., & Smeltzer, C. H. The Scientific construction of an interviewing chart. Personnel, 1936, 13, 14-19

Adams, E. L., Jr. Use of student reaction questionnaires at service schools. Military Review, 1950, 29(11), 58-62.

Questionnaire Theory and Development	17
Psych. Abst., 24, # 4350 (Rev. from rept.)	R-H

Adams, H.E., & Kirby, A. C. Manifest anxiety, social desirability, or response set. Journal of Consulting Psychology, 1963, 27(1), 59-61.

Response Bias, Question Stem	12, 3g
Psych. Abst., 37, #8006	A-H

Adams, H. F. An objectivity-subjectivity ratio for scales of measurement. Journal of Social Psychology, 1930, 1, 122-135.

Adams, H. F. Validity, reliability, and objectivity. Psychological Monographs: General and Applied, 1936, 47, 329-350.

Adams, J. F. An evaluation of the effect of level of item difficulty on various indices of item-discrimination. Dissertation Abstracts, 1959, 20, 1066-1067.

* Adams, J. S. An experiment on question and response bias. Public Opinion Quarterly, 1956, 20, 593-598.

Response Alternatives, Response Bias 3g, 12

ORA R-H

Adkins, D. C. A comparative study of methods of selecting items. Columbus, Ohio: Ohio State University Library, 1937.

N/A

N/A T-M

Adkins, D. C. Needed research on examining devices. American Psychologist, 1948, 3, 104-106.

Adkins, D. C. A commentary on multiple-choice item criteria. Public Personnel Review, 1958, 19, 296-298.

Achievement Measures 18

Psych. Abst., 33, # 11.072 A-N

The Advertising Research Foundation. Sources of Published Advertising Research. 1960.

The Advertising Research Foundation. Journal of advertising research cumulative index. (Index to all articles in the Journal from 1960 on.)

Ager, J., Reece, M., & Saltz, E. Studies of forced-choice methodology: Individual differences in social desirability. Educational and Psychological Measurement, 1962, 22, 365-370.

* Agostini, J. M. The case for direct questions on reading habits. *Journal of Advertising Research*, 1962, *4*(2), 28-33.

Semantic Differential Items, Validity	15
ORA	R-M

Ahlgren, A. *Reliability, predictive validity, and personality bias of confidence-weighted scores* (ERIC Document Reproduction Service, ED 033 384). Washington, D.C.: American Educational Research Association, 1968.

Achievement Measures	18
ERIC Document Reproduction Service, ED 033 384	A-N

Aiken, E. G. Alternate forms of a semantic differential for measurement of changes in self-description. *Psychological Reports*, 1965, *16*, 177-178.

Aiken, L. R., Jr. Frequency and intensity as psychometric response variables. *Psychological Reports*, 1962, *11*(2), 535-538.

Investigator Error, Response Alternatives	3g, 8
Psych. Abst., *37*, # 8007	A-H

Aikenhead, G. S. *A new methodology for test construction in course evaluation* (ERIC Document Reproduction Service, ED 080 312). Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Detroit, Michigan, 1973.

Achievement Measures	18
ERIC Document Reproduction Service, ED 080 312	A-N

Air University, Educational Services Division, Test Construction Unit. *Test construction and interpretation*. *Bulletin No. 1*. Maxwell Field, Ala.: Air University, 1946.

N/A	N/A
N/A	T-M

Ajzen, I., & Fishbein, M. The prediction of behavior from attitudinal and normative variables. Journal of Experimental Social Psychology, 1970, 6, 466-487.

Alderfer, C. P. Convergent and discriminant validation of satisfaction and desire measures by interviews and questionnaires. Journal of Applied Psychology, 1967, 51, 509-520.

Aleamoni, L. M. MERMAC: A model and system for instructional test and questionnaire analysis. Behavior Research Methods & Instrumentation, 1971, 3(4), 213-216.

* Alilunas, L. J. The personal setting as an influence in the study of the attitudes of college freshmen toward capitalism. School and Society, 1949, 69, 284-286.

Attitude Measures, Questionnaire Theory and Development 12, 13

Psych. Abst., 26, #1433 (Rev. from rept.) R-M

Alker, H. A., et al. Multiple-choice questions and student characteristics. Journal of Educational Psychology, 1969, 60(3), 231-243.

Allen, I. L. Detecting respondents who fake and confuse information about question areas on surveys. Journal of Applied Psychology, 1966, 50(6), 523-528.

Allen, I. L., & Colfax, J. D. Respondent's attitudes toward legitimate surveys in four cities. Journal of Marketing Research, 1968, 5, 431-433.

Allport, G. W., Vernon, P. E., & Lindzey, G. Study of values. Boston: Houghton Mifflin, 1960.

Alpert, M. I. Identification of determinant attributes: A comparison of methods. Journal of Marketing Research, 1971, 7, 184.

Question Stem, Validity, Bibliography, Instrument Format 3g, 15, 16

ORA R-H

* Altemeyer, R. A. Adverbs and intervals: A study of Likert scales. American Psychological Association, Proceedings of the Annual Convention of, 1970, 5(pt. 1), 397-398.

Adjectives, Rating Scales, Response Alternatives 2, 3d, 3f, 6

R-H

Alutto, J. A. Some dynamics of questionnaire completion and return among professional and magerial personnel: The relative impacts of reception at work site or place of residence. Journal of Applied Psychology, 1970, 54 (5), 430-432.

* Ambler, R. K., Blair, J. T., deRivera, J., Nelson, P. D., & Schoenberger, R. W. A note comparing the interview and written questionnaire techniques for identifying anxiety toward flying. USN School of Aviation Medical Research Report, Subtask 1, No. 20, Project No. NM 16 01 11, 1958.

Interviews, Personality Measures 1

Psych. Abst., 34, #2216 A-H

Anastasi, A. An empirical study of the applicability of sequential analysis to item selection. Educational and Psychological Measurement, 1953, 13, 3-13.

Question Stem 3g

Psych. Abst., 28, # 109

Anastasi, T. E., Jr. Face to face communication. Cambridge, Mass.: Management Center, 1967.

Anastasi, A. Psychological testing. (3d ed.) Macmillan, 1968.

Anderberg, M. R. Cluster analysis for applications. New York: Academic Press, 1972.

Anderson, D. H., & Petersen, D. F. Closing the communications gap with item sampling. Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, 1971.

* Anderson, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 1968, *9*, 272-278.

Adjectives 6

Journal of Personality and Social Psychology, *9*, p. 272 R-NA

Andrews, F. M., Morgan, J. N., & Sonquist, J. A. *Multiple classification analysis: A report on a computer program for multiple regression using categorical predictors*. Ann Arbor, Mich.: University of Michigan, Survey Research Center, Institute for Social Research, 1969.

Data Analysis 18

ORA R-N

Andrews, R. S. *A psychological technique for selection of personnel to perform as observer-recorders on field texts*. QM R&E Evaluation Agency, Technical Report R-3, Project No. 07-98-05-001, FEA MRS 4801 (MRS 48-7f), 1958.

N/A N/A

N/A T-H

Angoff, W. H. An empirical approach to a problem of psycholophysical scaling. *Journal of Applied Psychology*, 1949, *33*, 59-68.

Angoff, W. H. Test reliability and effective test length. *Psychometrika*, 1953, *18*, 1-14.

Angoff, W. H. *Basic equations in scaling and equating*. Statistical Report 61-51. Princeton, New Jersey: Educational Testing Service, 1961.

Annis, A. D., & Meier, N. C. The induction of opinion through suggestion by means of "planted content." *Journal of Social Psychology*, 1934, *5*, 65-81.

N/A 18

ORA R-N

Anonymous. Mogul "semantic differential" aims to provide qualitative research data. *Advertising Age*, 1958, 29(36), 3.

* Appel, V. An experimental test of the superiority and theory of forced-choice questionnaire construction. *Dissertation Abstracts*, Sept. 1959, 20, 1067-1068.

Forced-Choice Items, True-False Items, Validity, Instrument Length 2, 5

Dissertation Abstracts, 20, p. 1067 (Rev.) A-H

Arentsen, K. An investigation of the questionnaire method by means of the Cornell Index (Form N2). I. Review of the literature and method: Results for a group of military recruits. II. Results for a group of military medical patients. *Acta Psychiatrica et Neurologica, Kjøbenhavn*, 1957, 32(3), 231-279.

Military Personnel, Personality Measures 18

Psych. Abst., 33, # 1240 A-N

* Armstrong, J. S., & Overton, T. Brief vs. comprehensive descriptions in measuring intentions to purchase. *Journal of Marketing Research*, 1971, 8, 114-117.

Clarity 3g, 4

ORA R-H

Arnold, H. L. Analysis of discrepancies between true-false and simple recall examination. *Journal of Educational Psychology*, 1927, 14, 414-420.

Aronson, A. H. Service rating plans. *Public Personnel Review*, 1941, 2, 298-305.

Arthol, R. P., & Bridge, C. *Project Echo. Phase I*. (AD-657 613). Santa Barbara, California: General Research Corporation, June 1967. Report No. GRC-CR-0040-1.

N/A N/A

N/A T-M

Asch, M. J. Negative response bias and personality adjustment. Journal of Counseling Psychology, 1958, 5, 206-210.

Asch, S. E. Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego standards. Journal of Social Psychology, 1940, 12, 433-465.

Asch, S. E., Block, H., & Hertzman, M. Studies in the principles of judgments and attitudes: I. Two basic principles of Judgment. Journal of Psychology, 1938, 5, 219-251.

* Aschal, A. P. Relative values of poll-end and open-end questions in search for reasons of a problem. Educational Psychology, Delhi, 1958, 5, 55-60.

Open-Ended Items, Response Bias, Close-Ended Items	2, 12
Psych. Abst., 33, #8117 (Rev.)	A-M

* Ash, P., & Abramson, E. The effect of anonymity on attitude-questionnaire response. Journal of Abnormal and Social Psychology, 1952, 47(3), 722-723.

Anonymous Respondent, Rating Scales	7, 11
Psych. Abst., 27, #3444	A-H

Ashburn, R. An experiment in the essay-type question. Journal of Experimental Education, 1938, 7, 1-3.

Ashburn, R. R., & Bradshaw, J. H. An experiment in the continuity-type question. Journal of Educational Research, 1953, 47, 201-209.

Achievement Measures	18
Psych. Abst., 28, #6584	A-N

Asher, J. J. The Q by Q interview: With applications to personnel selection and to decision-making by courtroom juries (AD-711 066). San Jose, California: San Jose State College, 1967. Contract NONR-4817(00).

N/A	N/A
N/A	T-M

Asher, J. J. How the applicant's appearance affects the reliability and validity of the interview. Educational and Psychological Measurement, 1970, 30(3), 687-695.

Athey, K. R., Coleman, J. E., Reitman, A.P., & Tang, J. Two experiments showing the effect of the interviewer's racial background on responses to questionnaires concerning racial issues. Journal of Applied Psychology, 1960, 44(4), 244-246.

* Atkin, C. K., & Chaffee, S. H. Instrumental response strategies in opinion interviews. Public Opinion Quarterly, 1972, 36(1), 69-79.

Interviews, Response Bias, Investigator Error 12

ORA R-H

Attneave, F. A method of graded dichotomies for the scaling of judgments. Psychological Review, 1949, 56, 334-340.

Scaling, Paired Comparison Items, Ranking 14,2

ORA R-N

Auld, F., Jr. Influence of social class on personality test responses. Psychological Bulletin, 1952, 49, 318-332.

Ausubel, D. P., & Schpoont, S. H. Prediction of group opinion as a function of extremeness of predictor attitudes. Journal of Social Psychology, 1957, 46, 19-29.

Axelrod, J. N. Attitude measures that predict purchase. Journal of Advertising Research, 1968, 8(1), 3-17.

Rating Scales, Paired Comparison Items, Forced Choice Items, Card Sort, Open-ended Items, Ranking, Reliability, Validity 2

ORA R-H

Ayad, J. M., & Farnsworth, P.R. Shifts in the values of opinion items; further data. Journal of Psychology, 1953, 36, 295-298.

Azrin, N. H., Holz, W., & Goldamond, I. Response bias in questionnaire reports. *Journal of Consulting Psychology*, 1961, *25*, 324-326.

* Babatz, S. The effect of race of experimenter, instruction, and comparison population upon level of reported anxiety in negro subjects. *Journal of Personality and Social Psychology*, 1967, *7*(2), 194-196.

Investigator Error, Respondent's Motivation 9,11

Journal of Personality and Social Psychology, *7*, 194 R-M

Bachrack, S. D., & Scoble, H.M. Mail questionnaire efficiency: Controlled reduction of nonresponse. *Public Opinion Quarterly*, 1967, *31*, 265-271.

Back, K. W., & Gergen, K. J. Idea orientation and ingratiation in the interview: A dynamic model of response bias. *Proceedings of the Social Statistics Section*, American Statistical Association, 1963, 284-288.

* Back, K. W., Hill, R., & Stycos, J. M. Interviewer effect on scale reproducibility. *American Sociological Review*, 1955, *20*, 443-446.

Interviews. Investigator Error, Response Bias 10, 12

ORA R-M

* Baehr, M. E. A simplified procedure for the measurement of employee attitudes. *Journal of Applied Psychology*, 1953, *37*, 163-167.

Attitude Measures, Instrument Format, Response Alternatives, Scoring, Rating Scales 2, 3a, 3c, 8

ORA A-H

Baier, D. E. Reply to Travers' "A critical review of the validity and rationale of the forced-choice technique." *Psychological Bulletin*, 1951, *48*, 421-434.

Forced Choice Items, Questionnaire Theory and Development 14

ORA R-N

Bain, R. Theory and measurement of attitudes and opinions. Psychological Bulletin, 1930, 27, 357-379.

Bain, R. Stability in questionnaire response. American Journal of Sociology, 1931, 37(3), 445-453.

Baker, E. L. The effects of manipulated item writing constraints on the homogeneity of test items. Center for the Study of Evaluation Reprint Series No. 11. California University, Los Angeles: Center for the Study of Evaluation, 1970. Spons. Agency-Office of Education(DHEW), Washington D.C., Bureau of Research (BR-6-1646).

Achievement Measures 18

ERIC Document Reproduction Service, ED 036 870 A-N

Bakker, F. J. Emige probleme und methoden der itemanalyse. (Some problems and methods of item analysis.) Diagnostica, 1958, 4, 41-48.

Achievement Measures 18

Psych. Abst., 33, #9253 A-N

Baker, H. J. The construction and statistical interpretation of psychological tests. Review of Educational Research, 1932, 2, 295-299.

* Balinsky, B., Blum, M. L., & Dutka, S. The coefficient of agreement in determining product preferences. Journal of Applied Psychology, 1951, 35, 348-351.

Paired Comparison Items, Preference Measures, Reliability, Data Analysis 2, 8, 14

Psych. Abst., 26, #4227 (Rev. from rept.) R-M

* Ballin, M., & Farnsworth, P. R. A graphic rating method for determining the scale values of statements in measuring social attitudes. Journal of Social Psychology, 1941, 13, 323-327.

Rating Scales, Scaling 14, 2

ORA R-H

* Banta, T. J. Social attitudes and response styles. *Educational and Social Measurement*, 1961, *21*, 543-557.

Response Bias, Rating Scales — 12, 2

Psych. Abst., *36*, #4 GD 43 13 — A-M

Banta, T. J. Critical note on unidimensional tests. *Psychological Reports*, 1962, *11*(2), 449-450.

Response Bias — 14, 12

Psych. Abst., *37*, #7975 — A-N

Barclay, A., & Thumin, F. J. A modified semantic differential approach to attitudinal assessment. *Journal of Clinical Psychology*, 1963, *19*(3), 376-378.

Semantic Differential Items, Personality Measures — 2

Psych. Abst., *39*, #7574 — A-N

* Barclay, J. E., & Weaver, H. B. Comparative reliabilities and ease of construction of Thurstone and Likert attitude scales. *Journal of Social Psychology*, 1962, *58*(1), 109-120.

Scaling, Reliability — 2, 14

Psych. Abst., *37*, #6564 — A-M

* Barclay, W. D. The semantic differential as an index of brand attitude. *Journal of Advertising Research*, 1964, *4*(1), 30-33.

Semantic Differential Items, Validity — 2

ORA — R-H

Barker, W. S., & Gorham, W. A. *A research on the acceptance of QM clothing and equipment*. Psychological Research Associates, PRF Report 55-3, 1958.

N/A — N/A

N/A — T-H

* Barnes, T. I. The critical incident technique. *Sociology and Social Research*, 1960, *44*, 345-347.

Critical Incident Technique 17

ORA R-N

Barnett, L. D. Responses to questionnaires completed inside and outside the classroom. *The Family Life Coordinator*, 1965, *14*(3), 130-132.

Barnhart, E. N. A computational short cut in determining scale values for ranked items. *Psychometrika*, 1939, *4*, 241-242.

Barrett, F. D., Jr. The Canis method of reducing bias in survey research. *Journal of Marketing Research*, 1972, *9*, 329-330.

N/A 18

ORA A-NA

* Barrett, R. S., Taylor, E. K., Parker, J. W., & Martens, L. Rating Scale Content: I. Scale information and supervisory ratings. *Personnel Psychology*, 1958, *11*, 333-346.

Rating Scales, Response Alternatives, Investigator Error 3f, 12

Psych. Abst., *33*, #11. 121 A-H

Barron, F. Some personality correlates of independence of judgment. *Journal of Personality*, 1952, *21*, 387-297.

Barron, A. S. *The effects of three styles of interviewing on the response of women from two contrasting socioeconomic groups*. Ph.D. dissertation, Columbia University, 1957.

Barron, B. A., Gluckman, M., & Hirsch, J. The construction and calibration of behavioral rating scales. *Behavioral Science*, 1970, *15*(3), 220-226.

Reliability, Data Analysis 18

ORA R-NA

* Bartlett, C. J. Factors affecting forced-choice response. Personnel Psychology, 1960, 13, 399-406.

Forced Choice Items, Investigator Error 13

Psych. Abst., 36, #3HE99B A-M

* Bartlett, C. J., & Doorley, R. Social desirability response differences under research, simulated selection, and faking instructional sets. Personnel Psychology, 1967, 20(3), 281-228.

Investigator Error, Response Bias, Clarity 7, 11

ORA R-M

* Bartlett, C. J., Heermann, E., & Rettig, S. A comparison of six different scaling techniques. Journal of Social Psychology, 1960, 51, 343-348.

Paired Comparison Items, Ranking, Reliability, Rating Scales, Scaling 2

Psych. Abst., 34, #7620 A

Bartlett, C. J., Quay, L. C., and Wrightsman, L. S., Jr. A comparison of two methods of attitude measurement: Likert-type and forced choice. Educational and Psychological Measurement, 1960, 20, 699-704.

Forced-Choice Items, Response Bias, Rating Scales 2, 12

Psych. Abst., 35, #3338 A-M

* Bartlett, C. J., & Sharon, A. T. Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 1969, 22(3), 251-263.

Investigator Error, Raters, Rating Scales, Forced Choice Items 7, 13, 2

Personnel Psychology, 22, p. 262 (Rev. from rept.) R-M

Barton, Allen J. Asking the embarrassing question. Public Opinion Quarterly, 1958, 22, 67-68.

Question Stem	3g, 3c
Bureau of Census, #7111032601	A-N

Bashook, P. G., & Foster, S. F. How many ES are there? A critical analysis of problems concerning determination of evaluative factors of semantic differential scales. Paper presented at annual meeting of the American Educational Research Association, New Orleans, Louisiana, 1973. Pub. Date 73; Note-10 p.

Basic bibliography on marketing research, American Marketing Association, 1963.

Bass, B. M. Authoritatianism or acquiescence? Journal of Abnormal Psychology, 1955, 51, 616-623.

* Bass, B. M., & Wurster, C. R. Using the mark sense for ratings and personal data collection. Journal of Applied Psychology, 1956, 40(4), 269-271.

Scoring	8
ORA	R-M

* Bauer, E. J. Response bias in a mail survey. Public Opinion Quarterly, 1947, 11, 594-600.

Investigator Error	9
Potter, Sharpe, Hendee and Clark, 1972	A-N

* Bauer, R. K., & Meissner, F. Structure of mail questionnaires test of alternatives. Public Opinion Quarterly, 1963, 27(2), 307-311.

Instrument Format, Instrument Length	3g, 5
Potter, Sharpe, Hendee, & Clark, 1972 (Rev.)	R-H

Bauman, L. J., Rogers, T. F., & Weiss, C. H. *Abstracts of Papers On Respondent-Interviewer Interaction in the Research Interview*. Columbia University: Bureau of Applied Social Research, 1971.

Bibliography	17
ORA	T-H

Baumgarten, F. A proverb test for attitude measurement. *Personnel Psychology*, 1952, 5, 249-261.

Open-ended Items, Attitude Measures	18
Psych. Abst., 27, #6798	A-N

Baryoff, A. G., & Burke, J. H. The rater's guide. *Personnel Psychology*, 1950, 3, 461-465.

Military Personnel	17
Psych. Abst., 25, #3978	A-M

* Baryoff, A. G., Haggerty, H. R., & Rundquist, E. A. Validity of ratings as related to rating techniques and conditions. *Personnel Psychology*, 1954, 7, 93-112.

Raters, Forced Choice Items, Rating Scales, Validity, Military Personnel	1, 2, 7, 9
Psych. Abst., 29, #3086	R-H

* Becker, B. W., & Myers, J. G. Yeasaying response style. *Journal of Advertising Research*, 1970, 10(6), 31-36.

Response Bias, Semantic Differential Items	12
ORA	R-H

Becker, S. L. Why an order effect? *Public Opinion Quarterly*, 1954, 18, 271-278.

Check List, Instrument Format, Investigator Error, Response Alternatives, Instrument Length	13, 3a, 5, 3b
ORA	R-H

Beckwith, N. E., & Lehmann, D. R. The importance of differential weights in multiple attribute models of consumer attitude. Journal of Marketing Research, 1973, 10, 141-5.

Bejar, I. I., & Weiss, D. J. Comparison of four empirical differential item scoring procedures. Proceedings of the 81st Annual Convention of the American Psychological Association, Montreal, Canada, 1973, 8, 31-32.

Scoring 8

N/A T-N

Beldo, L. A. A multiple classification study of phrasing attitude statements. Dissertation Abstracts, 1953, 13, 124 6. (Abstract of Ph.D. thesis.)

* Belkin, M., & Lieberman, S. Effect of question wording on response distribution. Journal of Marketing Research, 1967, 4, 312-313.

Question Stem, Interviews 3g

ORA R-M

Bell, R. Q., Hartup, W. W., & Crowell, D. H. Mailed versus supervised administration of a projective questionnaire. Journal of Consulting Psychology, 1962, 26 (3), 290.

Bell, F. O., Hoff, A. L., & Hoyt, K. B. Answer sheets do make a difference. Personnel Psychology, 1964, 17(1), 65-71.

Instrument Format, Military Personnel, Scoring 3g

ORA R-M

Beloff, H. Two forms of social conformity: Acquiescence and conventionality. Journal of Abnormal and Social Psychology, 1958, 56, 99-104.

Bellows, R. M., & Estep, M. F. Employment psychology: The interview. New York: Holt, Rinehart and Winston, 1954.

Belson, W. A. Matching and prediction on the principle of biological classification. *Applied Statistics*, 1959. 65-75.

Belson, W. A. *Research methods: A bibliography of research papers relating to the adequacy of methods in survey research*. The Market Research Society, 1960.

N/A N/A

N/A T-M

Belson, W. A. The methods of research. *Business Review*, 1961.

Belson, W. A. Research on question design. *Business Review* (Australia), 1964, 7, 14-19.

* Belson, W. A. *A study of the effects of reversing the order of presentation of verbal rating scales*. London: Reprint Series, Survey Research Unit, London School of Economics and Political Science, 1965.

Response Alternatives, Rating Scales 3b, 9

ORA R-H

Belson, W. A. The effects of reversing the presentation order of verbal rating scales. *Journal of Advertising Research*, 1966, 6(4), 30-37.

Instrument Format, Literature Review, Rating Scales, Response Alternatives 3b

ORA R-H

Belson, W. A. Tape recording: Its effects on accuracy of response in survey interviews. *Journal of Marketing Research*, 1967, 4, 253.

Interviews 18

Bureau of Census - Survey Methodology Information System A-N

Belson, W. A. A volunteer bias in test-room groups. *Public Opinion Quarterly*, 1969, *24*(1), 115-126.

Respondent's Motivation, Response Bias, Investigator Error 11,12, 9

ORA R-M

* Belson, W. A. *Interviewer deviation from instructions in telling respondents how to use semantic differential scales*. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science, n.d. (a).

Semantic Differential Items, Investigator Error 7

ORA R-H

* Belson, W. A. *Research into question design*. London: Reprint Series Survey Research Unit, London School of Economics and Political Science, n.d. (b).

Investigator Error, Questionnaire Theory and Development, Literature Review 16, 3g

ORA R-M

* Belson, W. A. *Respondent understanding of survey questions*. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science, n.d. (c).

Clarity, Interviews, Investigator Error 4, 13

ORA R-H

Belson, W. A. *Studies in Readership*, London: Business Publications Ltd., 1962.

N/A N/A

N/A T-M

Belson, W. A., & Bell, C. R. *A bibliography of papers bearing on the adequacy of techniques used in survey research*. London: Oakwood, 1960 (a).

Belson, W., & Bell, C. R. *A bibliography: Techniques used in survey research*. The Market Research Society, 1960(b).

* Belson, W. A., & Duncan, J. A. A comparison of the check-list and open-response questionnaire systems. *Applied Statistics*, 1962, *11*,(3), 120-132.

Check List, Open-Ended Items, Response Alternatives 2

ORA R-H

Belson, W. A., & Thompson, B. *Bibliography of methods of social and business research*. London: London School of Economics and Political Science and Crosby Lockwood, 1973.

Belson, W. A., & Yule, V. R. *The semantic differential scaling system in market research. II. Accuracy of ratings*. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science.

Clarity, Semantic Differential Items, Reliability, Investigator Error, Response Alternatives 4, 7, 3f

ORA R-H

* Bendig, A. W. The reliability of self-ratings as a function of the amount of verbal anchoring and the number of categories on the scale. *Journal of Applied Psychology*, 1953, *37*, 38-41.

Instrument Format, Rating Scales, Response Alternatives, Reliability 3a, 3f, 8

Journal of Applied Psychology, 37, pp. 40-41. R-H

* Bendig, A. W. Reliability and the number of rating scale categories. *Journal of Applied Psychology*, 1954 (a), *38*, 38-40.

Preference Measures, Reliability, Rating Scales, Response Alternatives 3a, 8, 13

Psych. Abst., *29*, #81 (Rev. from rept.)

* Bendig, A. W. Transmitted information and the length of rating scales. Journal of Experimental Psychology, 1954 (c), 47, 303-308.

Response Alternatives, Rating Scales 3a, 8, 3g

Journal of Experimental Psychology, 47, p. 307 R-M

Bendig, A. W. Rater reliability and the heterogeneity of the scale anchors. Journal of Applied Psychology, 1955, 39, 37-39.

Investigator Error, Preference Measures, Raters, Response Bias, Response Alternatives, Question Stem 3f, 12, 13, 3g

Psych. Abst., 30, #68 (Rev.)

Bendig, A. W. "Social desirability" and "anxiety" variables in the IPAT Anxiety Scale. Journal of Consulting Psychology, 1959, 23, 377.

* Bendig, A. W., & Hughes, J. B. Effect of amount of verbal anchoring and number of rating-scale categories upon transmitted information. Journal of Experimental Psychology, 1953, 46, 87-90.

Instrument Format, Rating Scales, Response Alternatives 3a, 3f, 8

Psych. Abst., 28, #5944 (Rev. from rept.) R-H

Benjamin, K. Combining responses on two forms of a questionnaire with options in inverse order. Public Opinion Quarterly, 1949-50, 13, 688-690.

Bennett, A. S. Some aspects of preparing questionnaires. Journal of Marketing, 1945, 10, 175-179.

Bennett, A. S. Observations on the so-called cheater problem among field interviewers. International Journal of Opinion and Attitude Research, 1948, 2(1), 70-84.

* Bennett, E. M., Alpert, R., & Goldstein, A. C. Communications through limited-response questioning. Public Opinion Quarterly, 1954, 18, 303-308.

Close-Ended Items, Interviews, Attitude Measures,
Investigator Error 1

ORA R-H

Bennett, E.M., Blomquist, R. L., & Goldstein, A. C. Response stability in limited-response questioning. Public Opinion Quarterly, 1954, 18, 218-233.

Bennett, J. F., & Hays, W. L. Multidimensional unfolding: Determining the dimensionality of ranked preference data. Psychometrika, 1960, 25, 27-43.

Bennett, P. D. (Ed.) Research methodology: Longitudinal analysis. Paper presented at the Proceedings of Fall Conference, American Marketing Association, Marketing & Economic Development, 1965, 205-275.

Benney, M. Riesman, D., & Star, S. A. Age and sex in the interview. American Journal of Sociology, 1956, 62, 143-152.

Benson, A. H. Paired comparison approach to evaluating interview. Journal of Marketing Research, 1969, 6, 66.

Benson, L. E. Studies in secret-ballot technique. Public Opinion Quarterly, 1941, 5, 79-82.

Benson, P. H. A paired comparison approach to evaluating interviewer performance. Journal of Marketing Research, 1969, 6(1), 66-70.

Benson, P. H. How many scales and how many categories shall we use in consumer research? - A comment. Journal of Marketing, 1971, 35(4), 59-61.

Ranking, Rating Scales, Response Alternatives 2, 3a, 3f

ORA R-N

Benson, R. M. *Effects of instructions and verbal modeling on health information reporting in household interviews*. Ann Arbor, Mich.: University of Michigan, Survey Research Center, PHS HS00624-02, 1973.

Bent, R. K. Various techniques of combining ratings. *Journal of Educational Psychology*, 1937, *28*, 65-70.

Data Analysis	15
ORA	R-N

* Bentler, P. M. Semantic space is (approximately) bipolar. *Journal of Psychology*, 1969, *71*, 33-40.

Adjectives, Semantic Differential Items	14
Journal of Psychology, *71*, p.39-40 (Rev. from rept.)	R-H

Bentler, P. M., & Jackson, D. N. Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 1971, *76*(3), 186-204.

N/A	N/A
N/A	T-M

Bentler, P. M., & Lauoie, A. L. A nonverbal semantic differential. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*(4), 491-496.

Clarity	4
ORA	R-N

Benton, A. L. Influence of incentives upon intelligence test scores of school children. *Journal of Genetic Psychology*, 1936, *49*, 494-496.

Respondent's Motivation	11
ORA	R-N

Benton, A. L., & Kornhauser, G. I. A study of "score faking" on a mechanical interest test. *Journal of the Association of American Medical Colleges*, 1948, *23*, 57-60.

Berlioz, L. Liaisons d'item. (Relationship of items.) Bulletin d'Etudes et Recherches Psychologique, 1962, 11(2), 117-134.

Data Analysis 8

Psych. Abst., 38, #6042 A-N

* Bernard, J. An experimental comparison of ranking and paired comparisons as methods of evaluating questionnaire items. Papers of the American Sociological Society, 1933, 28, 81-84.

Ranking, Paired Comparison Items 2

ORA R-H

Bernberg, R. E. The direction of perception technique of attitude measurement. International Journal of Opinion Attitude Research, 1951, 5, 397-406.

Questionnaire Theory and Development 18

Psych. Abst., 27, #336 A-N

* Bernhardson, C. S., & Fisher, R. J. The relationship between personal desirability and endorsement with a forced-choice technique. Psychological Abstracts, 1971, 46, #9073.

Personality Measures, Response Bias 10, 12

Psych. Abst., 46, #9073 A-H

Bernstein, L. A. Statistics for decisions. New York: Grosset & Dunlap, 1965.

Berrien, F. K. The sensitivity of employee attitude questionnaires (AD 226 671). New Brunswick, New York: Rutgers State University, 1959, Report No. TR-5.

Best, W. H. Some new directions in personnel appraisal. *Personnel*, 1957, *34*, 45-50.

Rating Scales	2
Psych. Abst., *33*, #2229	A-N

Betts, G. L. Test calibration for categorical classification. *Educational and Psychological Measurement*, 1949, *9*, 269-279.

Scaling	13
Psych. Abst., *26*, #2739	A-N

* Bevan, W., & Avant, L. I. Response latency, response uncertainty, information transmitted and the number of available judgmental categories. *Journal of Experimental Psychology*, 1968, *76*, 394-397.

Response Alternatives, Respondent's Motivation, Instrument Length	3a, 5
Journal of Experimental Psychology, *76*, pp. 394-397	

Bevis, J. C. Interviewing with tape recorders. *Public Opinion Quarterly*, 1949-50, *13*(4), 629-634.

Bhatt, I. D. Performance as a function of varying types of incentives and stresses. *Education and Psychology Review*, 1965, *5*, 167-173.

Bickman, L., & Henchy, T. (Eds.) *Beyond the laboratory: Field research in social psychology*. New York: McGraw-Hill, 1972.

Biddle, B. J. An application of social expectation theory to the initial interview. *Dissertation Abstracts*, 1958, *19*, 186.

Bingham, W. E., Jr. A study of the relations which the galvanic skin response and sensory reference bear to judgments of the meaningfulness, significance, and importance of 72 words. *Journal of Psychology*, 1943, *16*, 21-34.

Bingham, W. E., Jr. A study of the effect of the presence of the examiner upon test scores in individual testing. Journal of Applied Psychology, 1944, 28, 471-476.

* Bingham, W. V. Halo, invalid and valid. Journal of Applied Psychology, 1939, 23, 221-228.

Investigator Error, Raters 7, 12

Journal of Applied Psychology, 23, p.221 (Rev.) R-H

* Bittner, R. H., & Rundquist, E. A. The rank-comparison rating method. Journal of Applied Psychology, 1950, 34, 171-177.

Ranking, Paired Comparison Items, Rating Scales 2, 14

Psych. Abst., 25, #3979 A-M

Blake, R. H. A comparison of the test-retest reliability of picture and verbal forms of occupational interest inventories. Dissertation Abstracts, 1967, 27(9-A), 2868.

* Blake, R. Comparative reliability of picture form and verbal form interest inventories. Journal of Applied Psychology, 1969, 53(1), 42-44.

Interest Measures, Clarity, Reliability 8, 4

Journal of Applied Psychology, 53, 42 (Rev. from rept.) R-H

Blake, R., & Dennis, W. The development of stereotypes concerning the Negro. Journal of Abnormal and Social Psychology, 1943, 38, 525-531.

Blalock, H. M., Jr., & Blalock, A. G. (Eds.) Methodology in social research. New York: McGraw-Hill, 1968.

* Blankenship, A. B. Does the question form influence public opinion poll results? Journal of Applied Psychology, 1940(a), 24, 27-30.

Clarity, Investigator Error, Question Stem 4, 12, 3g

ORA R-H

Blankenship, A. B. The effect of the interviewer upon the responses in a public opinion poll. *Journal of Consulting Psychology*, 1940(b), *4*(4), 134-136.

* Blankenship, A. B. The influence of the question form upon the response in a public opinion poll. *Psychological Records*, 1940(c), *3*, 345-422.

Attitude Measures, Clarity, Question Stem, Validity 4, 3g

ORA R-H

Blankenship, A. B. Pre-testing a questionnaire for a public opinion poll. *Sociometry*, 1940(d), *3*, 263-269.

Blankenship, A. B. The "sample" study in opinion research. *Sociometry*, 1940(e), *3*, 271-275.

* Blankenship, A. B. Psychological difficulties in measuring consumer preference. *Journal of Marketing*, 1942, *6*, 66-75.

Questionnaire Theory and Development 14, 17

ORA R-M

Blankenship, A. B. *Consumer and opinion research; the questionnaire technique*. New York: Harper, 1943.

Blankenship, A. B. The choice of words in poll questions. *Sociology and Social Research*, 1949(a), *25*, 12-18.

Blankenship, A. B. Source of interviewer bias. *International Journal of Opinion and Attitude Research*, 1949(b), *3*(1), 95-98.

Blankenship, A. B. Let's bury paired comparisons. *Journal of Advertising Research*, 1966, *6*(1), 13-17.

Paired Comparison Items, Validity, Questionnaire Theory and Development 2, 14

ORA R-N

Blanz, F., Ghiselli, E. E. The mixed standard scale: A new rating system. *Personnel Psychology*, 1972, *25*, 184-189.

Blankenship, A. B., et al. Questionnaire preparation and interviewer technique. *Journal of Marketing*, 1949, *14*, 399-433.

Blass, T., Pope, B., & Siegman, A. W. Verbal indices of interpersonal imbalance in the interview. *Proceedings of the Annual Convention of the American Psychological Association*, 1970, *5*(Pt. 2), 525-526.

Bloch, E. L., Goodstein, L. D., Jourard, S. M., & Jaffe, P. E. Comment on "influence of an interviewer's disclosure on the self-disclosing behavior of interviewees." *Journal of Consulting Psychology*, 1971, *18*(6), 595-600.

* Block, J. A comparison of the forced and unforced Q-sorting procedures. *Educational and Psychological Measurement*, 1956, *16*, 481-493.

Card Sorts	2
Psych. Abst., *32*, #31	A-H

* Block, J. A comparison between ipsative and normative ratings of personality. *Journal of Abnormal and Social Psychology*, 1957, *54*, 50-54.

Questionnaire Theory and Development	14
Psych. Abst., *33*, #806	A-H

* Block, J. An unprofitable application of the semantic differential. *Journal of Consulting Psychology*, 1958, *22*, 235-236.

Check List, Semantic Differential	2
Psych. Abst., *35*, #4780	A-M

Block, J. *The challenge of response sets*. New York: Appelton-Century-Crofts, 1965.

Response Bias	12
The challenge of response sets. (rev.)	R-N

Block, J. On further conjectures regarding acquiescence. Psychological Bulletin, 1971, 76(3), 205-210.

* Bloxom, B. Effects of anger-arousing instructions on personality questionnaire performance. Educational and Psychological Measurement, 1968, 28(3), 735-745.

Personality Measures, Investigator Error 13

ORA R-M

* Blumberg, H. H., De Soto, C. B., & Kuethe, J. L. Evaluation of rating scale formats. Personnel Psychology, 1966, 19(3), 243-259.

Instrument Format, Raters, Rating Scales, Response Alternatives 2, 3b, 3c, 3a, 3e

Psych. Abst., 41, #00060 (Rev. from rept.) R-H

Boardus, E. S. Measuring social distance. Journal of Applied Sociology, 1925, 9, 299-308.

Bocher, W. Specific motivational and situational influences on the results of questionnaire studies. In Merz, F. (Ed.), Bericht uber den 25. Kongress der deutschen Gesellschaft fur Psychologie Munster, 1966, pp. 519-525.

Bock, D. G. The impact of rating errors on the use of rating scales in selected experiments in oral communication research. Dissertation Abstracts International, 1970, 30(10-A), 4584.

Bock, R., Dicken, C., & Van Pelt, J. Methodological implications of content-acquiescence correlations in the MMPI. Psychological Bulletin, 1969, 71, 127-39.

Bock, D., & Wood, R. Test theory. Annual Review of Psychology, 1971, 22, 193-224.

Achievement Measures, Literature Review 16

ORA R-N

Bodi, M. J. Statement scaling study for 14 rating characteristics of the Commander's evaluation report. Indianapolis, Indiana: U.S. Army Enlisted Evaluation Center, 1967. Technical Report No. 89.

Military Personnel, Rating Scales 14

Psych. Abst., 43, #7432 (Rev.) A-M

Boek, W. E., & Lade, J. H. A test of the usefulness of the postcard technique in a mail questionnaire study. Public Opinion Quarterly, 1963, 27, 303-306.

Bogart, L. No opinion, don't know, and maybe no answer. Public Opinion Quarterly, 1967, 31, 332.

Bohrstedt, G. W. A quick method of determining the reliability and validity of multi-item scales. American Sociology Review, 1969, 542-548.

Boisen, M. Special nonresponse tabulations of three items on long form schedules in mail return and nonmail return sample. Washington: Cleveland Special Census, Bureau of the Census, Cleveland Special Census Results Memorandum No. 52, (67-211 (MRD)), 1967.

Boldt, R. F. An approximately reproducing scoring scheme that aligns random response and omission (ERIC Document Reproduction Service, ED 057 074). Princeton, New Jersey: Educational Testing Service, 1971.

Achievement Measures 18

ERIC Document Reproduction Service, ED 057 074 A-N

Booker, H. S., & David, S. T. Differences in results obtained by experienced and inexperienced interviewers. London: London School of Economics and Political Science.

Borg, W. R., & Gall, M. D. Educational research: An introduction. New York: David McKay, 1971.

Borgatta, E. F. (Ed.) Sociological methodology. San Francisco: Jossey-Bass, 1969.

Borgatta, E. F., & Glass, D. C. Personality concomitant of extreme response set. Journal of Social Psychology, 1961, 55, 213-221.

Personality Measures, Response Bias 18

Journal of Social Psychology, 55, p. 220 R-NA

Bornstein, H., Jensen, B. T., Goldstein, L. G., Dunn, T. F., & Berkhouse, R. G. Evaluation of the basic military performance test. USA TAGO Personnel Research Branch, Technical Research Note No. 75, 1957.

Achievement Measures 18

Psych. Abst., 33, #2252 A-N

Borslow, B. The Edwards Personal Preference Schedule (EPPA) and fallability. Journal of Applied Psychology, 1958, 42, 22-27.

Borus, M. E. Response error and questioning technique in surveys of earning information. Journal of the American Statistical Association, 1970, 65(330), 566-575.

Bose, P. K. Some criteria in item selection techniques. Indian Journal of Psychology, 1958, 33, 101-107.

Achievement Measures 18

Psych. Abst., 35, #3948 A-N

Bottrill, J. An investigation of some intrinsic variables affecting test responses. Journal of Psychology, 1969, 71, 83-88.

Boulger, J. G. A comparison of two methods of obtaining factual and subjective life history data in follow-up studies: Structured interview vs. questionnaire. Dissertation Abstracts International, 1969, 30(2-B), 1882-1883.

Interviews, Validity, Close-ended items 1, 10

ORA A-M

* Boulger, J. G. Comparison of two methods of obtaining life history data: Structured interview versus questionnaire. Proceedings of the Annual Convention of the American Psychological Association, 1970, 5(Pt. 2), 557-558.

Interviews 1

Psych. Abst., 44, #18811 A-M

Bowen, J. H. Familiarity scale values for 420 nouns in twelve combinations of frequency of occurrence and conceptual categorization. Psychological Reports, 1969, 25(3), 899-910.

Questionnaire Theory and Development 14

ORA R-M

Bowers, D., & Fine, B. Core questionnaire format and positive response bias. Ann Arbor, Mich.: University of Michigan, Institute for Social Research, 1968. Research Bulletin No. 5.

N/A N/A

N/A T-H

Boyd, H. W., Jr., & Westfall, R. Interviewers as a source of error in surveys. Journal of Marketing, 1955, 19(4), 311-324.

* Boyd, H. W., Jr., & Westfall, R. Interviewer bias once more revisited. Journal of Marketing Research, 1965(a), 7, 249-253.

Investigator Error, Interviews, Literature Review 9, 12, 6

Bureau of Census, #7111023901 R-H

Boyd, H. W., Jr., & Westfall, R. Interviewer bias revisited. *Journal of Marketing Research*, 1965(b), *7*(1), 58-63.

Boyd, J. L., Jr., & Shimberg, B. *Handbook of performance testing: A practical guide for test makers*. Princeton, N. J.: Educational Testing Service, 1971.

Achievement Measures 18

ERIC Document Reproduction Service, ED 052 220 A-M

Boyd, R. M. Influence of test form on assessment. *Australian Journal of Education*, 1971, *15*(2), 161-170.

* Bradburn, N. M. Selecting the questions to be asked in surveys. *Monthly Labor Review*, 1970, *93*, 27-29.

Question Stem 3g

ORA R-M

* Bradburn, N. M., & Mason, W. M. The effect of question order on responses. *Journal of Marketing Research*, 1964, 57-61.

Instrument Format, Instrument Length, Literature Review, Open-ended Items, Interviews 3c

ORA R-H

Bradley, U. I., & McClelland, J. N. *Basic statistical concepts*: Palo Alto, Calif.: Scott, Foreman, 1963.

Brauer, R. L. *Preferences of occupants in military family housing* (DAOF 8131). Champaign, Illinois, OCE Construction Engineering Research Laboratory, 1973.

Questionnaire Theory and Development 18

DDC, #DAOF8131 T-N

* Braun, J. R. Forced-choice self-report devices: A look at some unwarranted claims. *Measurement and Evaluation in Guidance*, 1969, 2(3), 153-155.

Forced-Choice Items, Investigator Error	2, 13
ORA	R-H

* Brennan, R. D. Trading stamps as an incentive in mail surveys. *Journal of Marketing*, 1958, *22*(1), 306-307.

Respondent's Motivation	11
ORA	R-M

* Brenner, M. H. Test difficulty, reliability, and discrimination as functions of item difficulty order. *Journal of Applied Psychology*, 1964, *48*(2), 98-100.

Instrument Format, Achievement Measures, Reliability	3c
Psych. Abst., *39*, #5039	A-N

Bressler, J. Judgment in absolute units as a psychophysical method. *Archives of Psychology*, New York, 1933, No. 152.

Bricker, P. D. The identification of redundant stimulus patterns. *Journal of Experimental Psychology*, 1955, *49*, 73-81.

Brigham, C. A. *A study of error*. New York: College Entrance Examination Board, 1932.

Brigham, J. C., & Cook, S. W. The influence of attitude on judgments of plausibility: A replication and extension. *Educational and Psychological Measurement*, 1970, *30*(2), 283-292.

* Brinkmeier, I. H. Minor studies on objective examination methods. IV. Sentence-length as a specific determiner in true-false examinations. *Journal of Educational Research*, 1930, *22*, 203.

Investigator Error, True-False Items, Achievement Measures, Question Stem — 3g, 14

ORA — R-M

Brinkmeier, I. H., & Keys, N. Circumstantiality as a factor in guessing on true-false examinations. *Journal of Educational Psychology*, 1930, *21*, 681-694.

Brinkmeier, I. H., and Ruch, G. M. Minor studies on objective examination methods. III. Specific determiners in true-false statements. *Journal of Educational Research*, 1930, *22*, 110-118.

Brinton, J. E. Deriving an attitude scale from semantic differential data. *Public Opinion Quarterly*, 1961, *25*(2), 289-295.

Britt, S. H. Why it's best to use combination research. *Printers' Ink*, 1954, *249*(4), 60-66.

Broadbent, D. E. Word-frequency effect and response bias. *Psychological Review*, 1967, *74*(1), 1-15.

Adjectives, Response Bias — 6, 12

ORA — R-N

* Brod, D., Kernoff, P., & Tewillinger, R. F. Anxiety and semantic differential responses. *Journal of Abnormal and Social Psychology*, 1964, *68*(5), 570-574.

Semantic Differential Items, Response Bias, Personality Measures — 12

Psych. Abst., *39*, #4647 — A-M

Broedling, L. *Development of attitudinal assessment techniques* (DN242001). Washington, D. C.: Naval Personnel R & D Laboratory, 1973 (a).

N/A	N/A
N/A	R-NA

Broedling, L. *Development of a small-sample methodology applicable to surveys of Navy personnel* (DN242017). Washington, D. C.: Naval Personnel R & D Laboratory, 1973 (b).

N/A	N/A
N/A	R-NA

Broen, W. E., Jr., & Wirt, R. D. Varieties of response sets. *Journal of Consulting Psychology*, 1958, *22*(3), 237-240.

Brogden, H. E. Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. *Psychometrika*, 1946, *11*, 197-214.

* Brokaw, L. D. Comparative validites of "short" versus "long" tests. *Journal of Applied Psychology*, 1951, *35*, 325-330.

Validity, Reliability, Instrument Length, Military Personnel	5
Psych. Abst., *26*, #4197	A-M

Brotemarkle, R. A., & Fernberger, S. W. A method for investigating the validity of the categories of a judgment test. *Journal of Educational Psychology*, 1934, *25*, 579-584.

Response Alternatives, Data Analysis, Psychophysical Measures, Validity, Scaling	3a, 8, 14
ORA	R-M

Brown, E. M. Influence of training, method, and relationship of the halo effect. Journal of Applied Psychology, 1968, 52(3), 195-199.

Brown, D. T. Stimulus similarity and the anchoring of subjective scales. American Journal of Psychology, 1953, 66, 199-214.

Response Alternatives, Psychophysical Measures 3f

ORA R-H

Brown, J. M. Respondents rate public opinion interviewers. Journal of Applied Psychology, 1955, 39(2), 96-102.

Brown, M. L. Use of a postcard query in mail surveys. Public Opinion Quarterly, 1965, 29, 635-637.

Brown, R. V. Evaluation of total survey error. Journal of Marketing Research, 1967, 4(2), 117-127.

Brown, R. W. Is a boulder sweet or sour? Contemporary Psychology, 1958, 3, 113-115.

Semantic Differential Items, Questionnaire Theory and Development 14

ORA R-N

Brown, S. R. Bibliography on Q technique and its methodology, Perceptual and Motor Skills, 1968, 26, 587-613.

* Brown, S. R. The forced-free distinction in Q technique. Journal of Educational Measurement, 1971, 8(4), 283-287.

Forced-Choice Items, Card Sorts 2, 18

Journal of Educational Measurement, 8, p. 286 R-H

Brown, W. P. Individual differences in associating to neutral and emotional words. Journal of Consulting and Clinical Psychology, 1970, 34, 33-36.

Adjectives, Personality Measures 6

Journal of Consulting and Clinical Psychology, 34, p.33 R-NA

Browning, R. C., et al. A study of officer rating methodology: X: Effect of selecter rater characteristics on validity of ratings. Report No. 909, 1952.

N/A N/A

N/A T-M

Brownless, V. T., & Keats, J. A. A retest method of studying partial knowledge and other factors influencing item response. Psychometrika, 1958, 23, 67-73.

Brueckner, L. J., & Hawkinson, M. J. The optimum order of arrangement of items in a diagnostic test. Elementary School Journal, 1934, 34, 351-357.

Brumback, G. B. A note on criterion contamination in the validation of biographical data. Educational and Psychological Measurement, 1969, 29 (2), 439-443.

* Brunner, G. A., & Carroll, S. J. The effect of prior notification on the refusal rate in fixed address surveys. Journal of Advertising Research, 1969, 9(1), 42-44.

Respondent's Motivation, Interviews 9, 11

ORA R-M

Brunswick, E. Systematic and representative design of psychological experiments. Berkeley & Los Angeles, Calif.: University of California Press, 1947.

Bruvold, W. H. Category and successive intervals scales for rating statements and stimulus objects. *Journal of Experimental Psychology*, 1969 *81*(2), 230-234.

Questionnaire Theory and Development, Scaling 14

ORA R-N

* Bruvold, W. H. Rater's attitudes and the method of equal-appearing intervals. *Proceedings of the Annual Convention of the American Psychological Association*, 1971, *6*(Pt. 1), 373-374.

Attitude Measures, Rating Scales, Response Bias 10, 12

ORA R-N

Bruvold, W. H. Consistency among attitudes, beliefs and behavior. *Journal of Social Psychology*, 1972, *86*, 127-134.

Buchanan, P. C., & Wiley, L. N. The use of a multi-purpose rank-rating scale for personnel evaluation and preference research. *American Psychologist*, 1948, *3*, 345.

Rating Scales, Ranking 2

American Psychologist, 3, p. 345 A-M

Buchwald, A. M. Variations in the apparent effects of "right" and "wrong" on subsequent behavior. *Journal of Verbal Learning and Verbal Behavior*, 1962, *1*(1), 71-78.

Response Alternatives, Investigator Error, Respondent's Motivation 3a, 11

Psych. Abst., *37*, #6128 A-M

* Bucklow, M. Staff reporting: A new look at the halo effect. *Personnel Practice Bulletin*, 1960, *16*(4), 29-33.

Investigator Error 13

Psych. Abst., *35*, #4045 (Rev.) A-N

Budd, W. C., & Blakely, L. S. Response bias in the Minnesota Teacher Attitude Inventory. *Journal of Educational Research*, 1958, *51*, 707-709.

Response Bias 12

Psych. Abst., *33*, #11043 A-N

Bueker, K. Reactions to a questionnaire survey: Some sources of respondent resistance. *Journal of Psychiatric Nursing & Mental Health Services*, 1969, *7* (5), 215-221.

Buel, W. D. A simplification of Hay's method of recording paired comparisons. *Journal of Applied Psychology*, 1960, *44*, 347-348.

* Buel, W. D. Preference indices stability in forced-choice rating scale construction. *Journal of Industrial Psychology*, 1963, *1*(2), 55-58.

Forced-Choice Items, Check-List 2, 14

Psych. Abst., *38*, #10.152 A-H

Bues, H. W. The construction and validation of a scale to measure attitude toward any practice. *Purdue University Studies for Higher Education, XXVI*, 1934, *35*, 64-67.

Bureau of the Census. *List of books with opinions and/or information relevant to questionnaire design*. Washington, D. C.: Bureau of the Census, RRB Report No. 63-52, (64-415 (MRD)), 1963.

Bureau of the Census. *Questionnaire research reports -- Series A, Report 3: Results of three classroom experiments with census of population housing questionnaires*. Washington, D. C.: Bureau of the Census, U. S. Department of Commerce, Social and Economic Statistics Administration, Statistical Res. Div. Response Research Staff Report No. 72-14.

Buros, O. (Ed.) *The mental measurement year books*. Highland Park, New Jersey: The Gryphon Press.

Burros, R. H. The estimation of the discriminal dispersion in the method of successive intervals. *Psychometrika*, 1955, *20*, 299-306.

Burton, A. C. *On the science of field testing as applied to clothing and personal equipment*. Defense Research Board, 1948.

* Burtt, H. E., & Gaskill, H. V. Suggestibility and the form of the question. *Journal of Applied Psychology*, 1932, *16*(4), 358-373.

Clarity, Question Stem 4, 3g

Journal of Applied Psychology, *16*, 373 (Rev. from rept.) R-H

Burwen, L. S., Campbell, D. T., & Kidd, J. The use of a sentence completion test in measuring attitudes toward superiors and subordinates. *Journal of Applied Psychology*, 1956, *40*, 248-250.

Bush, L. E. Individual differences multidimensional scaling of adjectives denoting feelings. *Journal of Personality and Social Psychology*, 1973, *25*(1), 50-57.

Clarity, Adjectives 4,9

Journal of Personality and Social Psychology, *25*(1), p. 50 R-H

* Buss, A. H. The effect of item style on social desirability and frequency of endorsement. *Journal of Consulting Psychology*, 1959, *23*, 510-513.

Response Bias, Question Stem 12, 3g

Journal of Consulting Psychology, 23, p. 512 R-H

* Cahalan, D. Effectiveness of a mail questionnaire technique in the army. *Public Opinion Quarterly*, 1951, *15*, 575-578.

Respondent's Motivation, Interviews 1, 11

Potter, Sharpe, Hendee and Clark, 1972 A-M

* Cahalan, D., Tamulonis, V., & Verner, H. M. Interviewer bias involved in certain types of opinion survey questions. *International Journal of Opinion and Attitude Research*, 1947, *1*(1), 63-77.

Investigator Error, Response Alternatives, True-False Items, Open-ended Items, Multiple Choice Items — 2, 12

Bureau of Census (7110001701) — A-H

Caldwell, L. S. *Military performance - Physical decrement and enhancement* (DA 086071). Fort Knox, Kentucky: Medical Research Laboratory, June 1973.

N/A — N/A

N/A — R-NA

Calhoun, R. L. Item form and item discriminating power: An experimental study. *Dissertation Abstracts*, 1962, *23*(1), 335-336.

Multiple Choice Items, Question Stem, Achievement Measures — 3g

Dissertation Abstracts, *23*(1), pp. 335-336(Rev.) — A-N

California State Department of Education, Bureau of School Planning. *Profile rating wheel: An instrument to evaluate school facilities. Revised edition* (ERIC Document Reproduction Service, ED 072 552). Sacramento: California State Department of Education, Bureau of School Planning, 1972.

Data Analysis — 8

ERIC Document Reproduction Service, ED 072 552 — A-N

Campbell, A. A. Two problems in the use of the open question. *Journal of Abnormal and Social Psychology*, 1945, *40*(3), 340-343.

Campbell, A. C. Some determinants of the difficulty of non-verbal classification items. *Educational and Psychological Measurement*, 1961, *21*, 899-913.

Campbell, D. P. *Some desirable characteristics of interest inventories* (ERIC Document Reproduction Service, ED 033 417). Paper presented at meeting of American Personnel and Guidance Association, Las Vegas, Nevada, 1969.

Campbell, D. T. The indirect assessment of social attitudes. *Psychological Bulletin*, 1950, *47* (1), 15-18.

Campbell, D. T., & Giske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81-105.

Campbell, D. T., Lewis, N. A., & Hunt, W. A. Context effects with judgmental language that is absolute, extensive, and extraexperimentally anchored. *Journal of Experimental Psychology*, 1958, *55*, 220-228.

Response Alternatives, Raters 3a, 3f

ORA R-N

Campbell, D. T., & Mohr, P. J. The effect of ordinal position upon responses to items in a check list. *Journal of Applied Psychology*, 1950, *34*, 62-67.

Check List, Response Alternatives 3b

ORA R-H

Campbell, J. T., & Rundquist, E. A. Scale items for inclusion in forced-choice rating forms. *American Psychologist*, 1950, *5*, 280.

* Campbell, D. T., Siegman, C. R., & Rees, M. B. Direction-of-wording effects in the relationship between scales. *Psychological Bulletin*, 1967, *68*, 293-303.

Instrument Format, Response Bias 3g, 12

ORA R-N

Campion, J. E., Gelfand, N. I., & Lassoff, S. Z. Biasing influence of interviewer expectations on interviewee responses in a marketing study. *Catalog of Selected Documents in Psychology*, 1972, *2*, 46-47.

Cannell, C., & Fowler, F. Comparison of a self-enumerative procedures and a personal interview: A validity study. *Public Opinion Quarterly*, 1963, *27*, 250-264.

Cannell, C. G., Fowler, F. J., & Marquis, K. H. The influence of interviewer and respondent psychological and behavioral variables on the reporting in household interviews. *Vital & Health Statistics*, 1968, Series 2.

Cannell, C. F., & Kahn, R. The collection of data by interviewing. In Festinger, L., & Katz, D. (Eds.), Research methods in the behavioral sciences, pp. 327-380. New York: The Dryden Press, 1953.

Cannon, D., & Olson, H. C. SPANOCON: Span of Control. II. Effect on reliability of free and forced distributions in rating. HumRRO Research Memorandum, Subtask Spanocon, Task 11-28, 1961.

Data Analysis	8
Psych. Abst., 37, #5705	A-N

Canter, F. M., & Canter, A. N. Authoritarian attitudes and adjustment in a military situation. United States Armed Forces Medical Journal, 1957, 8 1201-1207.

* Canter, R. R., Jr. A rating-scoring method for free-response data. Journal of Applied Psychology, 1953, 37, 455-457.

Scoring	8
Psych. Abst., 29, #71	A-M

Cantrell, G. K. Multiple-choice item form in relation to age, intelligence, and level of education. Brooks AFB, Texas: USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), October 1969. SAM-TR-69-65.

N/A	N/A
N/A	T-H

* Cantril, H. Experiments in the wording of questions. Public Opinion Quarterly, 1940(a), 4, 330-332.

Attitude Measures, Clarity, Investigator Error	4, 13, 3g
Potter, Sharpe, Hendee & Clark, 1972 (Rev.)	R-M

Cantril, H. Problems and techniques: Experiments in the wording of questions. Public Opinion Quarterly, 1940 (b), 4(2), 330-332.

Cantril, E. Gauging public opinion. Princeton, N. J.: Princeton University Press, 1947.

Interviews, Textbook 13, 17

Bureau of Applied Social Research A-H

Cantril, H., & Fried, E. The meaning of questions. In E. Cantrol (Ed.), Gauging Public Opinion. Princeton, N.J.: Princeton University Press, 1944.

Capron, V. L. Relative effect of three orders of arrangement of items upon pupils' scores in certain arithmetic and spelling tests. Journal of Educational Psychology, 1933, 24, 687-694.

Carlson, R. E. Effect of interview information in altering valid impressions. Journal of Applied Psychology, 1971, 55(1), 66-72.

Carpenter, J. B., & Dudycha, A. L. Effects of item format on item discrimination and difficulty. Journal of Applied Psychology, 1973, 58(1), 116.

Carper, J., & Doob, L. W. Intervening responses between questions and answers in attitude surveys. Public Opinion Quarterly, 1953-54, 17, 511-519.

Carroll, S. J., & Nash, A. N. Effectiveness of a forced-choice reference check. Personnel Administration, 1972, 35(2), 42-46.

Carter, H. D. Importance and significance of objective test items. California Journal of Educational Research, 1955, 6, 61-71.

Achievement Measures 18

Psych. Abst., 29, #8992 A-N

* Carter, R. F., Ruggels, W. L., & Chaffee, S. H. The semantic differential in opinion measurement. Public Opinion Quarterly, 1968, 32(4), 666-674.

Semantic Differential Items, Response Alternatives 3f, 14

ORA R-H

Cartwright, D. Relation of decision-time to the categories of response. American Journal of Psychology, 1941, 54, 174-196.

Response Alternatives, Clarity 3a, 4

ORA R-M

Casanova, T. The measurement of randomness in test items. Journal of Experimental Education, 1944 (a), 12, 196-183.

Casanova, T. The use of the method of runs for testing the randomness of the order of examination items. Journal of Experimental Education, 1944 (b), 12, 165-168.

* Cataldo, E. F., Johnson, R. M., & Kellstedt, L. A. Card sorting as a technique for survey interviewing. Public Opinion Quarterly, 1970, 34(2), 202-215.

Card Sorts, Response Bias 2, 12

Bureau of Census, #7122200202 A-H

* Cattell, R. B. Psychological measurement: ipsative, normative, and interactive. Psychological Review, 1944, 51, 292-303.

Questionnaire Theory and Development, Scaling 2, 14

ORA R-N

Cattell, R. B. Factor analysis. New York: Harper, 1952.

Cattell, R. B., & Saunders, D. R. Inter-relationship and matching of personality factors from behavior ratings, questionnaire and objective test data. Journal of Social Psychology, 1950, 31, 243-260.

* Cavan, R. S. The questionnaire in a sociological research project. American Journal of Sociology, 1933, 38(5), 721-727.

Question Stem 3g

Potter, Sharpe, Hendee and Clark, 1972 A-M

Central Statistical Board. *A report of the Central Statistical Board on the returns made by the public to the Federal Government*, House Doc. No. 27, 76th Congress, 1st Session, Washington, D. C., 1939.

N/A N/A

N/A T-N

* Champion, D. J., and Sear, A. M. Questionnaire response rate: A methodological analysis. *Social Forces*, 1969, *47*, 335-339.

Instrument Length, Respondents' Motivation 5, 11

ORA R-H

Champney, H., & Marshall, H. Optimal refinement of the rating scale. *Journal of Applied Psychology*, 1939, *22*, 323-331.

Rating Scales, Response Alternatives 3a

ORA R-M

Chapanis, A. *Research techniques in human engineering*. Baltimore: The Johns Hopkins Press, 1962.

Chapman, L. J., & Bock, R. D. Components of variance due to acquiescence and content in the F scale measure of authoritarianism. *Psychological Bulletin*, 1958, *55*, 328-333.

Chapman, L. J., & Campbell, D. T. Response set in the F scale. *Journal of Abnormal and Social Psychology*, 1957, *54*, 129-132.

Chapman, L. J., & Campbell, D. T. The effect of acquiescence response-set upon relationships among the F scale, ethnocentrism, and intelligence. *Sociometry*, 1959, *22*, 153-161.

Chein, R, An introduction of sampling, Appendix B in C. In Selltiz, M.J., Deutsch, M., and Cook, S. W. (Eds.) *Research methods in social relations*. New York; Holt, 1959.

Chevan, A. Minimum-error scalogram analysis. Public Opinion Quarterly, 1972, 36(3), 379-387.

Child, I. L. The use of interview data in qualifying the individual's role in the group. Journal of Abnormal and Social Psychology, 1943, 38, 305-318.

Choo, T. Communicator-credibility and communication-discrepancy as determinants of opinion change. Dissertation Abstracts, 1960, 21, 1246.

Choppin, B. H., & Purvis, A. C. A comparison of open-ended and multiple-choice items dealing with literacy understanding. Research in the Teaching of English, 1969, 3(1), 15-24.

Christal, R. E., & Madden, J. M. Effect of degree of familiarity in job evaluation. USAF WADD Personnel Laboratory Technical Note No. 60-143, 1960.

N/A N/A

N/A T-H

Christie, R., Havel, J., & Seidenberg, B. Is the F-Scale irreversible? Journal of Abnormal and Social Psychology, 1958, 56, 143-159.

* Clancy, K. J., & Garson, R. Why some scales predict better. Journal of Advertising Research, 1970, 10(5), 33-38.

Paired Comparison Items, Rating Scales, Response Bias 2, 9, 12

ORA R-H

* Clark, E. L. General response patterns to five-choice items. Journal of Educational Psychology, 1956, 47, 110-117.

Multiple Choice Items, Response Bias, Response Alternatives 3a, 12, 3b, 7

Psych. Abst., 31, #8813 (Rev. from rept.) R-H

Clark, E. L. Item difficulties based on end segments. Journal of Educational Psychology, 1957, 48, 457-459.

Achievement Measures	18
Psych. Abst., 33, #2024	A-N

* Clark, K. E., & Kreidt, P. H. An application of Guttman's new scaling techniques to an attitude questionnaire. Educational and Psychological Measurement, 1948, 8, 215-223.

Scaling, Attitude Measures	2, 14
Psych. Abst., 23, #3993	A-M

Clark, R. A. The use of the Q-sort for collecting attitude data from company commanders under field conditions. Paper for Western Psychological Association Meeting, Berkeley, Calif., March 1956.

N/A	N/A
N/A	T-M

Clarke, M. A. Arabic distractors for English vocabulary tests. English Language Teaching, 1972, 27(1), 73-76.

* Clausen, J. A., & Ford, R. N. Controlling bias in mail questionnaires. Journal of American Statistical Association, 1947, 42, 497-511.

Respondent's Motivation	11
ORA	R-M

Cleary, T. A., & Hilton, T. L. An investigation of item bias. American Psychologist, 1964, 19, 506.

* Clemans, W. V. An analytical and empirical examination of some properties of ipsative measure. Psychometrika, Monogram Supplement 14, 1965.

Scoring, Questionnaire Theory and Development	14
ORA	R-M

Cliff, N. *The relation of adverb-adjective word combinations to their components*. Princeton, New Jersey: Educational Testing Service, 1955. Research Memorandum 55-9.

N/A	N/A
N/A	T-M

* Cliff, N. Adverbs as multipliers. *Psychological Review,* 1959, *66*, 27-44.

Adjectives	6
ORA	R-M

Cliff, N. *Multidimensional scaling and cognition: I. Relations among judgments obtained under varying conditions*. Los Angeles, Ca.: University of Southern California, 1965. Technical Report No. 1.

N/A	N/A
N/A	T-M

Cliff, N. *Multidimensional scaling and cognition: II. The relation of evaluation to multidimensional meaning spaces*. Los Angeles, Ca.: University of Southern California, 1966 (a). Technical Report No. 2.

N/A	N/A
N/A	T-M

Cliff, N. *Multidimensional scaling and cognition: III. Threat evaluation and subjective organization of simulated raids*. Los Angeles, Ca.: University of Southern California, 1966 (b). Technical Report No. 3.

N/A	N/A
N/A	T-M

Cliff, N. *Multidimensional scaling and cognition: Final report*. Los Angeles, Ca.: University of Southern California, 1967. Final Report.

N/A	N/A
N/A	T-M

* Cliff, N. Adjective check list responses and individual differences in perceived meaning. *Educational and Psychological Measurement*, 1968 (a) *28* (2), 1063-1077.

Response Bias, Personality Measures, Check List	12
Educational and Psychological Measurement, *28*(2), p. 1076 (Rev. from rept.)	R-N

Cliff, N. The "idealized individual" interpretation of individual differences in multidimensional scaling. *Psychometrika*, 1968 (b), 33, 225-232.

Data Analysis, Questionnaire Theory and Development	15
Psychometrika, *33*, p. 225	R-N

Cliff, N. Liking judgments and multidimensional scaling. *Educational and Psychological Measurement*, 1969, *29*(1), 87-98.

Scaling	14
ORA	R-M

Cliff, N. Scaling. *Annual Review of Psychology*, 1973, *24*, 473-506.

Cliff, N., Pennell, R., & Young, F. W. Multidimensional scaling in the study of set. *American Psychologist*, 1966, *21*, 707.

Instrument Format, Investigator Error, Question Stem	3c, 12
American Psychologist, *21*, p. 707	R-M

* Cloud, J., & Vaughn, G. M. Using balanced scales to control acquiescence. Sociometry, 1970, 33, 193-202.

Attitude Measures, Response Bias, Scoring, Question Stem 12, 8, 3g

Bureau of Census, #7111038602 (Rev. from Rept.) R-H

Coble, J. A. Results of the pilot study racial match experiment. In Lansing, J. B., Withy, S. B., Wolfe, A. C., et al. (Eds.) Working papers on survey research in poverty areas. Ann Arbor, Mich.: University of Michigan, Institute for Social Research, Survey Research Center, 1971.

Cohen, N. E. The relativity of absolute judgments. American Journal of Psycnology, 1937, 49, 93-100.

Coffin, T. E. Some conditions of suggestion and suggestibility: A study of some attitudinal and situational factors influencing the process of suggestion. Psychological Monographs: General and Applied, 1941, No. 241.

Coffman, W. E. Estimating the internal consistency of a test when items are scored 2, 1, or 0. Educational and Psychological Measurement, 1952, 12, 392-393.

Data Analysis 18

Psych. Abst., 27, #5515 A-N

* Cohen, L. Use of paired comparison analysis to increase statistical power of ranked data. Journal of Marketing Research, 1967, 4, 509.

Data Analysis, Paired Comparison Items, Ranking 2, 8

ORA R-N

* Cohen, R. The position effects problem. Public Opinion Quarterly, 1965, 29, 456.

Instrument Format, Interviews 3c

Public Opinion Quarterly, 1965, 29(Fall), 456. A-H

* Cohn, T. S. The relation of the F Scale to a response set to answer positively. *Journal of Social Psychology*, 1956, *44*, 129-133.

Response Bias 12

Psych. Abst., *33*, #4684 A-N

Cohn, T. S. The relation of the F scale to intelligence. *Journal of Social Psychology*, 1957, *46*, 207-217.

Collet, L. S. Elimination scoring: An empirical evaluation. *Journal of Educational Measurement*, 1971, *8*(3), 209-214.

Scoring, Achievement Measures 18

ERIC Document Reproduction Service, EJ 144 176 A-N

Collet, L. S. *Elimination vs. best responses made for M-C tests* (ERIC Document Reproduction Service, ED 041 037). Paper presented at the joint annual meeting of American Educational Research Association and National Council on Measurement in Education, Minneapolis, Minn., 1970.

Achievement Measures 18

ERIC Document Reproduction Service, ED 041 037 A-N

Collins, G. On methods. *Journal of Advertising Research*, 1961, *1*(3), 28-33.

Paired Comparison Items, Ranking, Rating Scales, Scaling 2, 17

ORA R-N

Collins, W. A. *Idiosyncratic verbal behavior of interviewers*. Palo Alto, Ca.: Stanford University, Institute for Communication Research, October, 1968.

Collins, W. A. Interviewers' verbal idiosyncracies as a source of bias. *Public Opinion Quarterly*, 1970, *34*(1), 416-422.

* Colombotos, J., Elinson, J., & Loewenstein, R. Effect of Interviewers' sex on interview responses. *Milbank Memorial Fund Quarterly*, 1969, *47*(1, pt. 2) 227-232.

Interviews 9

Milbank Memorial Fund Quarterly, *47*, p. 230 R-M
(Rev. from rept.)

Connelly, G. M. Now let's look at the real problem: Validity. *Public Opinion Quarterly*, 1945, *9*, 51-60.

Connelly, G. M., & Harris, N. A symposium on interviewing problems. *International Journal of Opinion and Attitude Research*, 1948, *2*.

Cook, D. L. The use of free response data in writing choice-type items. *Journal of Experimental Education*, 1958, *27*, 125-133.

Multiple Choice Items, Achievement Measures, Open-Ended Items 14, 18

Psych. Abst., *34*, #2360 A-N

Cook, S. W. *A multiple-indicator approach to attitude measurement*. Boulder, Colo.: Colorado University, Institute of Behavioral Science, July 1963. Contract AFOSR436 63.

Cook, S. W. *Studies of attitude and attitude measurement*. Boulder, Colo.: Colorado University, Institute of Behavioral Science, March 1968. Project AF-9778.

Cook, S. W., & Selltiz, C. A multiple-indicator approach to attitude measurement. *Psychological Bulletin*, 1964, *62*(1), 36-55.

* Coombs, C. H. Psychological scaling without a unit of measurement. *Psychological Review*, 1950, *57*, 145-158.

Scaling 14

ORA R-H

Coombs, C. H. *A theory of psychological scaling*. Ann Arbor, Mich.: University of Michigan Press, 1952.

Coombs, C. H., Milhalland, J. E., & Womer, F.B. The assessment of partial knowledge. *Educational and Psychological Measurement*, 1956, *16*, 13-17.

* Cooper, A., & Cowen, E. L. The social desirability of trait descriptive terms: A study of feeling reactions to adjective descriptions. *Journal of* Social Psychology, 1962, *56*, 207-215.

Adjectives, Rating Scale, Response Bias 6, 12

Psych. Abst., *37*, #1223 A-H

Copeland, H. A. Studies in the reliability of personnel records. *Journal of Applied Psychology*, 1968, *22*, 247-251.

Corah, N. L., Feldman, M. J., Cohen, I. S., Gruen, W., Meadow, A., & Ringwall, E. A. Social desirability as a variable in the Edwards Personal Preference Schedule. *Journal of Consulting Psychology*, 1958, *22*, 70-72.

Response Bias, Interest Measures 12

Psych. Abst., *33*, #5978 A-N

Corey, L. G. How to isolate product attributes. *Journal of Advertising Research*, 1970, *10*(4), 41-44.

Data Analysis 8

ORA R-N

* Corey, S. M. Signed versus unsigned attitude questionnaires. Journal of Educational Psychology, 1937, 28(2), 144-148.

Anonymous Respondent	11
Potter, Sharpe, Hendee, and Clark, 1972	A-H

Costin, F. The optimal number of alternatives in multiple-choice achievement tests: Some empirical evidence for a mathematical proof. Educational and Psychological Measurement, 1970, 30(2), 353-8.

Costin, F. Three-choice versus four-choice items: implications for reliability and validity of objective achievement tests. Educational and Psychological Measurement, 1972, 32(4), 1035-8.

Coston, M. L., Thorne, H. W., & York, C. M. Study of bias in new equipment training. Fort Benning, Georgia: U.S. Army Infantry Board, April, 1973. TECOM Project No. 9-CO-00F-000-010, USAIB Project No. 3381.

Cotton, J. W. Elementary statistical theory for behavior scientists. Palo Alto, Calif: Addison-Wesley, 1967.

* Couch, A., & Keniston, K. Yeasayers and naysayers: Agreeing response set as a personality variable. Journal of Abnormal and Social Psychology, 1960, 60, 151-174.

Attitude Measures, Response Bias	2, 12
Psych. Abst., 34, #7376	R-H

Coward, A. F., & Lord, F. M. Test reliability as a function of the number of choices per item. Research Bulletin 50-47. Princeton, New Jersey: Educational Testing Service, 1950.

Cox, W. E. Response patterns to mail surveys. Journal of Marketing Research, 1966, 3, 392-7.

Cragun, J. R., & McCormick, E. J. *Job inventory information: Task and scale reliabilities and scale interrelationships*. USAF PRL Technical Report, No. 67-15, 1967.

* Crawford, P. L. Comparison of two attitude scaling methods. *Psychological Reports*, 1965, 17(3), 681-682.

Rating Scales, Paired Comparison Items 2

Psych. Abst., 40, #3522 A-H

Creelman, M. B. *The experimental investigation of meaning: A review of the literature*. New York: Springer, 1966.

N/A N/A

N/A T-H

Crehan, K. D., & Slakter, M. J. Note on comparison of paired-multiple-response items and multiple-choice items. *Psychological Reports*, 1971, 28 (1), 310.

Crespi, I. Use of a scaling technique in surveys. *Journal of Marketing*, 1961, 25 (5), 69-72.

Crocker, L. M., & Mehrens, W. A. *The comparative effectiveness of different item analysis techniques in increasing change score reliability*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, 1971.

Cronbach, L. H. Studies of acquiescence as a factor in the true-false tests. *Journal of Educational Psychology*, 1942, 33, 401-415.

* Cronbach, L. J. An experimental comparison of the multiple true-false and multiple choice. *Journal of Educational Psychology*, 1941(a), 32, 533-543.

Multiple Choice Items, True-False Items, Achievement Measures 2

Journal of Educational Psychology, 32, p. 561 R-H

Cronbach, L. J. The true-false test: A reply to Count Etoxinod. *Education*, 1941 (b), *42*, 59-61.

* Cronbach, L. J. Response set and test validity. *Educational and Psychological Measurement*, 1946, *6*, 475-494.

Response Bias, Multiple Choice Items 12

ORA R-H

* Cronbach, L. J. Further evidence on response sets and test design. *Educational and Psychological Measurement*, 1950, *10*, 3-31.

Response Bias 12

ORA R-H

Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, *16*, 297-334.

Cronbach, L. J. *Essentials of psychological testing*. (3rd ed.) New York: Harper & Row, 1969.

Textbook, Bibliography 15, 16, 17

ORA R-M

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley, 1972.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, *52*, 281-302.

Crosby, R. W. Attitude measurement in a bilingual culture. *Journal of Marketing Research*, 1969, *6*, 421.

Crosland, H. *The psychological methods of word association and reaction time as tests of deception*. University of Oregon Publication, 1929, 1, No. 1.

Cross, O. H. A study of faking on the Kuder Preference Record. *Educational and Psychological Measurement*, 1950, *10*, 271-277.

Crowne, D. P., & Marlowe, D. *The approval motive: Studies in evaluative dependence*. New York: John Wiley, 1964.

Crutchfield, R. S., & Gordon, D. A. Variations in respondents' interpretations of an opinion-poll question. *International Journal of Opinion and Attitude Research*, 1947, *1*(3), 1-12.

Attitude Measures, Clarity, Question Stem, Investigator Error	12, 4
Census Abstracts, 71110071	A-N

Cuber, J. F., & Gerberich, J. B. A note on consistency in questionnaire responses. *American Sociological Review*, 1946, *11*, 13-15.

Cummings, J. D. *Naval personnel attitudes toward environmental pollution* (AD 901 680L). San Diego, California: Navy Manpower and Material Analysis Center Pacific, 1971. Report No. WSR-72-5.

Attitude Measures	18
DDC	A-N

Curtis, F. D., Darling, W. C., & Shearman, N. H. A study of the relative values of two modifications of the true-false test. *Journal of Educational Research*, 1943, *36*, 517-527.

Curtis, H. A., & Kropp, R. P. *Experimental analyses of the effects of various modes of item presentation on the scores and factorial content of tests administered by visual and audiovisual means - A program of studies basic to television testing*. Tallahassee, Florida: Florida State University, School of Education, 1961. Report No. NDEA-VIIA-385.

Dailey, J. G., and Hagenah, T. *Vocational interest measurements, theory and practice*. University of Minnesota Press, 1955.

* Dakin, R. E., & Tennant, D. Consistency of response by event-recall intervals and characteristics of respondents. *Sociological Quarterly*, 1968, *9*, 73-84.

Investigator Error	10
Bureau of the Census, #7131907101	A-H

Dalkey, N. C. *Delphi* (AD-660 554). Santa Monica, California: Rand Corporation, 1967. Report No. P-3704.

N/A	N/A
N/A	T-M

Dalkey, N. C. *Experiments in group prediction* (AD 668 107). Santa Monica, California: Rand Corporation, 1968. Report No. P-3820.

Damarin, F., & Messick, S. J. *Response styles as personality variables: A theoretical integration of multivariate research*. Research Bulletin 65-10. Princeton, New Jersey: Educational Testing Service, 1965.

N/A	N/A
N/A	T-M

Danielson, W. A. A data reduction method for scaling dichotomous items. *Public Opinion Quarterly*, 1957, *21*, 377-379.

Darling, A. B., & Bragdon, H. W. Essay vs. objective testing in social studies. *College Board Review*, 1952, *16*, 260-263.

Multiple Choice Items, Open-ended Items	2
Psych. Abst., *27*, #663	A-N

Darnell, D. K. A technique for determining the evaluative discrimination capacity and polarity of semantic differential scales for specific concepts. *Dissertation Abstracts*, 1964, *25*(2), 2623-2624.

Davis, R. J. Considerations for evaluating evaluation instruments. *Ill. Teacher Contemp. Roles*, 1970, *13*(6), 282-284.

Dawes, R. M. *Fundamentals of attitude measurement*. New York: John Wiley, 1972.

Textbook	17
ORA	T-H

Dawis, R. V. *The measurement of employee attitudes* (AD 720 712). Minneapolis, Minnesota: Center for the Study of Organizational Performance and Human Effectiveness, 1971. Report No. TR-3001.

Dawis, R. V., & Tinsley, H. E. A. *The equivalence of semantic and figural test presentation of the same items*. Minneapolis, Minn.: Minnesota University Center for the Study of Organizational Performance and Human Effectiveness. Report No. TR-3004.

Day, R. L. Systematic paired comparisons in preference analysis. *Journal of Marketing Research*, 1965, *2*, 406.

Paired Comparison Items, Preference Measures	2
Journal of Marketing Research, 1965, p. 406 (Rev.)	A-M

Day, R. L. Methods of estimating consumer preference distributions. *California Management Review*, 1967, *9*, 35-42.

Dean, J. P., & Whyte, W. F. How do you know if the informant is telling the truth. *Human Organization*, 1958, *18*(4), 21-23.

DeFleur, M. L., & Catton, W. R., Jr. The limits of determinacy in attitude measurement. *Social Forces*, 1957, 35, 295-300.

DeGreene, K. *Systems psychology*. New York: McGraw-Hill, 1970.

Demb, G. Comparative of positive and negative multiple-choice test items. *University of North Carolina...Research in Progress. University of North Carolina Rec*., 1951, *60*(492), 251. (Abstract of Master's thesis.)

Deming, W. E. On errors in surveys. *American Sociological Review*, 1944, *9*, 359-369.

Deming, W. E. On the distinction between enumerative and analytic surveys. *Journal American Statistical Association*, 1953, *43*, 244-255.

* Deri, S., Dinnerstein, D., Harding, J., & Pepitone, A. D. Techniques for the diagnosis and measurement of intergroup attitudes and behavior. *Psychological Bulletin*, 1948, *45*, 248-271.

Attitude Measures	1
Psych. Abst., *23*, #700 (Rev. from rept.)	R-N

DeSoto, C. B. The predilection for single orderings. *Journal of Abnormal and Social Psychology*, 1961, *62*, 16-23.

Dexter, L. A. Role relationships and conceptions of neutrality in interviewing. *American Journal of Sociology*, 1956, *62*(2), 153-157.

Diamond, J. J., & Evans, W. J. An investigation of the cognitive correlates of test-wise-ness. *Journal of Educational Measurement*, 1972, *9*(2), 145-150.

* Diers, C. J. Social desirability and acquiescence in response to personality items. Dissertation Abstracts, 1961, 22(5), 1709.

Response Bias, Personality Measures 12

Dissertation Abstracts, 22, p. 1709 (Rev.) A-H

* Dillehay, R. C., & Jernigan, L. R. The biased questionnaire as an instrument of opinion change. Journal of Personality and Social Psychology, 1970, 15, 144-155.

Attitude Measures, Investigator Error 10

Journal of Personality and Social Psychology, 15, p. 144 (Rev.) R-H

Divesta, F. J., & Walls, R. T. Response selection as a function of instructions and motivation under nonreinforcement conditions. Journal of Experimental Psychology, 1967, 73(3), 365-373.

Paired Comparison Items, Respondent's Motivation 11

Journal of Experimental Psychology, 73, p. 365 R-NA

* Dixon, T. R. Experimenter approval, social desirability and statements of self-reference. Journal of Consulting and Clinical Psychology, 1970, 35, 400-405.

Respondent's Motivation, Response Bias 11, 12

Journal of Consulting and Clinical Psychology, 35, p. 400 (Rev. from rept.) R-N

* Dizney, H. F., Merrifield, P. R., & Davis, O. L., Jr. Effects of answer-sheet format on arithmetic test scores. Educational and Psychological Measurement, 1966, 26(2), 491-493.

Instrument Format, Achievement Measures 3g

Psych. Abst., 40, #13532 A-M

* Dodd, S. C., & Gerberick, T. R. Word scales for degrees of opinion. Language and Speech, 1960, 3, 18-31.

Response Alternatives, Adjectives	3b, 6
Psych. Abst., 35, #6315	R-H

* Dohrenwend, B. S. Some effects of open and closed questions on respondent's answers. Human Organization, 1965, 24(2), 175-184.

Closed-Ended Items, Open-Ended Items	1, 3g
ORA	R-H

Dohrenwend, B. S. An experimental study of directive interviewing. Public Opinion Quarterly, 1970, 34(1), 117-125.

Dohrenwend, B. S. An experimental study of payments to respondents. The Public Opinion Quarterly, Winter 1970-71, 34(4), 621-624.

N/A	N/A
N/A	T-H

Dohrenwend, B. S., Colombotos, J., & Dohrenwend, B. P. Social distance and interviewer effects. Public Opinion Quarterly, 1968, 32(3), 410-422.

Dohrenwend, B. S., & Richardson, S. A. A use for leading questions in research interviewing. Journal of Marketing, 1940, 5(2), 122-124.

Question Stem	3g
Bureau of Census, #7111006801	A-N

Dohrenwend, B. S., Williams, J. A., & Weiss, C. H. Interviewer biasing effects: Toward a reconciliation of findings. Public Opinion Quarterly, 1969, 33(1), 121-129.

Doll, R. E. Item susceptibility to attempted faking as related to item characteristic and adopted fake set (AD 740 290). San Diego, California: Navy Medical Neuropsychiatric Unit, 1970. Report No. NMNRU-69-33.

N/A N/A

N/A T-H

Dollard, J. Under what conditions do opinions predict behavior? Public Opinion Quarterly, 1948-49, 12, 623-632.

Dolliver, R. H., & Clark, J. A. Status faking on the SVIB-M. Journal of Vocational Behavior, 1972, 2(1), 47-55.

Donald, M. N. Implications of nonresponse for the interpretation of mail questionnaire data. Public Opinion Quarterly, 1960, 24, 99-114.

* Doncel, J. F., Alimena, B. S., & Birch, C. M. Influence of prestige suggestion on the answers of a personality inventory. Journal of Applied Psychology, 1949, 33, 352-355.

Personality Measures, Rating Scales, Response Bias 10, 12

ORA R-M

Downie, N. M. Fundamentals of measurement. (2nd ed.) Oxford University Press, 1967.

Draguns, J. G. Response sets on the MMPI and in structuring ambiguous stimuli. Psychological Reports, 1963, 13(3), 823-828.

Response Bias, Personality Measures 12

Psych. Abst., 38, #8484 (Rev.) A-N

Drayton, L. E. Bias arising in wording consumer questions. Journal of Marketing, 1954, 19, 140-145.

Dressel, P. L., & Schmid, J. Some modifications of the multiple-choice item. Educational and Psychological Measurement, 1953, 13, 574-595.

Drinkwater, B. L. A comparison of the direction-of-perception techniques with the Likert method in the measurement of attitudes. Journal of Social Psychology, 1965, 67(2), 189-196.

Response Alternatives, Scales, Reliability 2, 3f

Psych. Abst., 40, #4747 (Rev. from rept.) R-H

Driver, R. A. The validity and reliability of ratings. Personnel, 1941, 17, 185-191.

Druckman, D., & Dunning, M. C. Double agreement of reversed item attitude scales: Agreement set versus social desirability. Psychonomic Bulletin, 1967, 1(2), 19.

* Dubeck, J. A., et al. Falsification of the forced guilt inventory. Journal of Consulting and Clinical Psychology, 1971, 36(2), 296.

Response Bias 12

ERIC Document Reproduction Service, EJ 040 673 A-M

* Dubois, C. The card-sorting or psychophysical interview. Public Opinion Quarterly, 1949-50, 13(4), 619-628.

Investigator Error, Card Sorts, Response Bias, Respondent's Motivation 2, 11, 12, 13

ORA R-N

DuBois, P., Loeving, J., & Smith, T. L., Jr. Evaluation for methods of keying psychological tests for prediction of external criteria. USAF Personnel and Training Center, Research Reproduction No. AFPTCR-TN-56-65, 1956.

Question Stem 3g

Psych. Abst., 31, #6907 A-N

Dudycha, G. J. A note on the "halo effect" in ratings. *Journal of Social Psychology*, 1942, *15*, 331-333.

* Dunlap, J. W. Problems arising from the use of a separate answer sheet. *Journal of Psychology*, 1940, *10*, 3-48.

Instrument Format	2, 3e, 14, 3g
Report summary	R-H

Dunlap, J. W., DeMello, A., & Cureton, E. E. The effects of different directions and scoring methods on the reliability of a true-false test. *School and Society*, 1929, *30*, 378-382.

Dunn, T. F., Goldstein, L. G., & Berkhouse, R. G. *Effect of item construction principles on difficulty, reliability, and validity*. USA TAGO Personnel Research Branch, Technical Research Note 64, 1956.

Achievement Measures	18
Psych. Abst., *32*, #2779	A-N

Dunnette, M. D. Accuracy of students' reported honor point averages. *Journal of Applied Psychology*, 1952, *36*, 20-22.

* Dunnette, M. D., Aylward, M., & Uphoff, W. H. The effect of lack of information on the undecided response in attitude surveys. *Journal of Applied Psychology*, 1956, *40*(3), 150-153.

Response Alternatives, Clarity	3a, 4
ORA	R-H

* Dunnette, M. D., & Heneman, H. G., Jr. Influence of scale administrator on employee attitude responses. *Journal of Applied Psychology*, 1956, *40*, 73-77.

Anonymous Respondent, Investigator Error	3g, 11, 9
Journal of Applied Psychology, *40*, p. 77 (Rev. from rept.)	R-H

Dunn-Rankin, P. The true probability distribution of the range of rank totals and its application to psychological scaling. *Dissertation Abstracts*, 1966, *26*(8), 4827.

* Durant, H., & Maas, I. Who doesn't answer? *British Psychological and Sociological Bulletin*, 1956, *29*, 33-34.

Instrument Length, Respondent's Motivation 5, 11

Potter, Sharpe, Hendee and Clark, 1972 A-M

Durbin, J., & Stuart, A. *Differences in response rates of experienced and inexperienced interviewers*. London: London School of Economics and Political Science, Survey Research Centre. n.d.

* Dyer, R. F., Klein, R. D., & Yudowitch, K. L. *Analysis of alternate forms of a VOLAR/MVA questionnaire*. Palo Alto: Operations Research Associates, 1974. (Prepared for the Army Research Institute for the Behavioral and Social Sciences, Fort Hood, Texas under Contract DAHC19-74-C-0032.)

Question Stems, Response Alternatives 3b, 3c

ORA

* Dyer, R. F., Matthews, J. J., Wright, C. E. & Yudowitch, K. L. *Questionnaire construction manual*. Palo Alto: Operations Research Associates, 1975.

Questionnaire Theory and Development 14, 17

ORA R-H

Ebel, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, *16*, 407-424.

Ebel, R. L. *Measuring educational achievement*. New Jersey: Prentice-Hall,1965.

Ebel, R. L. Expected reliability as a function of choices per item. *Educational and Psychological Measurement*, 1969, *29*(3), 565-570.

Ebel, R. L. The test for true-false test items. *School Review*, 1970, *78* (3), 363-389.

Ebel, R. L. *The comparative effectiveness of true-false and multiple choice achievement test items* (ERIC Document Reproduction Service, ED 050 148). Paper presented at the Annual Meeting of the American Educational Research Association, New York, 1971 (a).

Achievement Measures 18

ERIC Document Reproduction Service, ED 050 148 A-N

Ebel, R. L. How to write true-false test items. *Educational and Psychological Measurement*, 1971 (b), *31*(2), 417-426.

Ebel, R. L. Why is a longer test usually a more reliable test? *Educational and Psychological Measurement*, 1972, *32*(2), 249-253.

Achievement Measures, Instrument Length 5

ERIC Document Reproduction Service, EJ 061 302 A-N

Ebel, R. L. Test development, interpretation, and use. *TM Report* *19*. Princeton, New Jersey: Educational Testing Service, 1973.

Literature Review 16

ORA R-H

Echternacht, G. J., et al. *An evaluation of the feasibility of confidence testing as a diagnostic aid in technical training* (ERIC Document Reproduction Service, ED 058 318). Princeton, New Jersey: Educational Testing Service, 1971.

Eckland, B. K. Effects of prodding to increase mail-back returns. *Journal of Applied Psychology*, 1965, *49* (3), 165-169.

Edgerton, H. A., Britt, S. H., & Norman, R. D. Objective differences among various types of residents to a mailed questionnaire. *American Sociological Review*, 1947, *12*, 435-444.

Edgerton, H. A., & Kolbe, L. E. The method of minimum variation for the combination of criteria. *Psychometrika*, 1936, *1*, 183-187.

* Edrich, H. The effects of context and of raters' attitudes on judgments of favorableness of statements about a social group (AD 653 916). Boulder, Colorado: University of Colorado, 1965. Project AF-9778. Institute of Behavioral Science.

Question Stem 3g

DDC, # AD 653 916 A-M

Educational Testing Service. Multiple choice questions: A close look (ERIC Document Reproduction Service, ED 081 783). Princeton, New Jersey: Educational Testing Service, 1973.

Achievement Measures 18

ERIC Document Reproduction Service, ED 081 783 A-N

Edwards, A. L. Assessing opinion and discovering facts. Psychology for the Armed Forces. Washington, D. C., Infantry Journal Press, 1945. Chapter 21.

Edwards, A. L. A comparison of the Thurstone and Likert methods of attitude scale construction. Journal of Applied Psychology, 1946 (a), 30(1), 72-83.

* Edwards, A. L. A critique of "neutral items" in attitude scales constructed by the method of equal-appearing intervals. Psychological Review, 1946 (b), 53, 159-169.

Reliability, Scaling 2, 8

ORA R-M

Edwards, A. L. On Guttman's scale analysis. Educational and Psychological Measurement, 1949, 8, 313-318.

Scaling 14

ORA R-H

* Edwards, A. L. Psychological scaling by means of successive intervals. Psychometric Laboratory: University of Chicago, 1951, 69, 1-13.

Scaling 14

ORA R-H

Edwards, A. L. The scaling of stimuli by the method of successive intervals. Journal of Applied Psychology, 1952, 36, 118-122.

Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. _Journal of Applied Psychology_, 1953, _37_(2), 90-93.

Edwards, A. L. Experiments: Their planning and execution. In Lindzey, G. (Ed.), _Handbook of social psychology_. Cambridge, Mass.: Addison-Wesley, 1954.

Reliability, Scaling 8, 2

ORA R-M

* Edwards, A. L. Social desirability and Q sorts. _Journal of Consulting Psychology_, 1955, _19_, 462.

Card Sorts, Response Bias 12

ORA R-H

* Edwards, A. L. A technique for increasing the reproducibility of cumulative attitude scale. _Journal of Applied Psychology_, 1956, _40_, 263-265.

Attitude Measures, Paired Comparison Items, Rating Scales, Scaling 2, 14

Psych. Abst., _31_, #5916 (Rev. from rept.) R-H

* Edwards, A. L. Social desirability and probability of endorsement of items in the interpersonal check list. _Journal of Abnormal and Social Psychology_, 1957 (a), _55_, 394-396.

Anonymous Respondent, Response Bias 11, 12

Journal of Abnormal and Social Psychology, _55_, p. 396 R-M

Edwards, A. L. _The social desirability variable in personality assessment and research_. New York: Dryden Press, 1957 (b).

Edwards, A. L. _Techniques of attitude scale construction_. New York: Appleton-Century-Crofts, 1957(c).

Edwards, A. L. Statistical analysis. (Rev. ed.) New York: Holt, Rhinehart, and Winston, 1960.

Edwards, A. L. Correlations between scores on personality scales when items are stated in the first and third person form. Educational and Psychological Measurement, 1969, 29(3), 561-563.

Edwards, A. L., & Abbott, R. D. Measurement of personality traits: Theory and technique. Annual Review of Psychology, 1973, 24, 241-278.

Literature Review 16

PASAR, 50, #03043 R-NA

Edwards, A. L., & Diers, C. J. Social desirability and the factorial interpretation of the MMPI. Educational and Psychological Measurement, 1962, 22, 501-509.

* Edwards, A. L., & Diers, C. J. Neutral items as a measure of acquiescence. Educational and Psychological Measurement, 1963, 23(4), 687-698.

Response Bias, Personality Measures 12

Psych. Abst., 38, #8486 A-H

Edwards, A. L., Diers, C. J., & Walker, J. N. Response sets and factor loadings on sixty-one personality scales. Journal of Applied Psychology, 1962, 46, 220-225.

* Edwards, A. L., & Horst, P. Social desirability as variable in Q technique studies. Educational and Psychological Measurement, 1953, 13, 620-625.

Response Bias, Card Sorts 12

Educational and Psychological Measurement, 13, p. 625 R-H

* Edwards, A. L., & Kenny, K. C. A comparison of the Thurstone and Lickert methods of attitude scale construction. Journal of Applied Psychology, 1946, 30, 72-83.

Attitude Measures, Reliability, Scaling 2, 8, 14

ORA R-H

Edwards, A. L., & Kilpatrick, F. P. Scale analysis and the measurement of social attitudes. Psychometrika, 1948 (a), 13, 99-114.

* Edwards, A. L., & Kilpatrick, F. P. A technique for the construction of attitude scales. Journal of Applied Psychology, 1948 (b), 32, 374-384.

Attitude Measures, Scaling 2, 14

ORA R-H

Edwards, A. L., & Thurstone, L. L. An internal consistence check for scale values determined by the method of successive intervals. Psychometrika, 1952, 17, 169-180.

Edwards, A. L., & Walsh, J. A. Response sets in standard and experimental personality scales. American Educational Research Journal, 1964, 1(1), 52-61.

* Ehrlich, J. S., & Reisman, D. Age and authority in the interview. Public Opinion Quarterly, 1961, 25(1), 39-56.

Investigator Error 9

Bauman, Rogers, and Weiss, 1971 (Rev.) A-M

Eisenberg, P. Two methods of combining attitudes of like, indifference, and dislike, into one score. Journal of Applied Psychology, 1945, 29, 246-251.

* Eisenberg, T. Fakability and validity of two types of forced-choice formats. Dissertation Abstracts, 1965, 26(3), 1790-1791.

Response Bias, Forced Choice Items 2, 12

Dissertation Abstracts, 26(3), pp. 1790-1791 (Rev.) A-H

Eisenman, R., & Rappaport, J. Complexity preference and semantic differential ratings of complexity-simplicity and symmetry-asymmetry. Psychonomic Science, 1967, 7(4), 147-148.

Eisler, H. The connection between magnitude and discrimination scales and direct and indirect scaling methods. Psychometrika, 1965, 30(3), 271-289.

Scaling, Response Alternatives, Data Analysis, Paired Comparison Items 8

Psych. Abst., 40, #2091 A-M

Ekehammar, B., & Magnusson, D. Subjective confidence and interjudge agreement as functions of amount of information: A study of interview data. Stockholm: Psychological Laboratories, Univ. of Stockholm, 1972, No. 366.

Ekman, G. Scales of Conservatism (AD 688 342). Stockholm, Sweden: Stockholm University, Psychological Laboratories, January 1963.

* Ekman, G., & Kunnapas, T. Note on direct and indirect scaling methods. Psychological Reports, 1960, 6, 174.

Rating Scales 2

Psych. Abst., 35, #5483 A-H

Ekman, G., & Sjoberg, L. Scaling. Annual Review of Psychology, 1965, 16, 451-474.

Elashoff, J. D., & Spiegel, D. E. Optimal choice of rater teams: II. Applications. *Psychometrika*, 1969, *34*(1), 33-44.

Elinson, J. Attitude research in the Army. *Journal of Applied Psychology*, 1949, *33*, 1-5.

Military Personnel, Attitude Measures 15

Psych. Abst., *23*, #3688 A-N

Elinson, J., & Cisin, I. H. Detection of interviewer cheating through scale technique. *Public Opinion Quarterly*, 1948, *12*.

Elison, J., & Haines, V. T. Role of anonymity in attitude surveys. *American Psychologist*, 1950, *5*, 315.

* Ellenbogen, B. L., & Danley, R. A. Comparability of responses to a socially concordant question: "Open-end" and "closed". *Journal of Health and Human Behavior*, 1962, *3*(2), 136-140.

Closed-Ended Items, Open-Ended Items 1, 2

Bureau of the Census, #7110004901 A-H

Elliot, F. R. Eye vs. ear in moulding opinion. *Public Opinion Quarterly*, 1936, *1*, 83-87.

Elliot, L. L. *Factorial structure of airman self-rating and their relationship to peer nominations* (AD 242 388). Wright Patterson AFB, Ohio: Aeronautical Systems Division, 1960.

N/A N/A

N/A R-NA

* Elliot, L. L. Effects of item construction and respondent attitude on response acquiescence. Educational and Psychological Measurement, 1961, 21, 405-415.

Response Bias, Question Stem, Military Personnel 12, 9

Psych. Abst., 36, #2H FOSE A-H

Ellis, A. Personality questionnaires. Review of Educational Research, 1947 (a), 17, 53-63.

* Ellis, A. Questionnaire versus interview methods in the study of human love relationships. American Sociological Review, 1947 (b), 12, 541-553.

Interviews 1

Psych. Abst., 23, #4196 A-H

* Ellis, A. Questionnaire versus interview methods in the study of human love relationships. II. Uncategorized responses. American Sociological Review, 1948, 13, 61-65.

Interviews 1

Potter, Sharpe, Hendee and Clark, 1972 A-M

Ellis, R. A., Endo, C. M., & Armer, M. The use of potential nonrespondents for studying nonresponse bias (AD-708 207). Eugene, Oregon: Oregon Univ., Dept. of Sociology, 1970. Contract No. AF-AFOSR-1582-68, PH-MH-15735.

N/A N/A

N/A T-M

Ellis, A., & Gerberich, J. R. Interests and attitudes. Review of Educational Research, 1947, 17, 64-77.

Elster, R. S., & Githens, W. H. Preferences for senior naval officers' billets obtained by using four different psychological scaling techniques. Proceedings of the Annual Convention of the American Psychological Association, 1972, 7(Pt. 1), 21-22.

Elster, R. S., & Githens, W. H. Development of a man-to-man rating scale for evaluating performance. *Proceedings of the 81st Annual Convention of the American Psychological Association*, Montreal, Canada, 1973, *8*, 761-762.

* Eng, E., & French, R. L. The determination of sociometric status. *Sociometry*, 1948, *11*, 368-371.

Paired Comparison Items, Rating Scales, Scaling	2, 8
Psych. Abst., *25*, #7384	A-H

Engel, J. F. Tape recorders in consumer research. *Journal of Marketing*, 1962, *26*(2), 73-74.

Interviews, Projective Items, Open-Ended Items	3g, 15
ORA	R-M

Engelhart, M. D. Suggestions for writing achievement exercises to be used in tests scored on the electric scoring machine. *Educational and Psychological Measurement*, 1947, *7*.

Engelhart, M. D. A comparison of several item discrimination indices. *Journal of Educational Measurement*, 1965, 2(1), 69-76.

Achievement Measures, Data Analysis	8
Psych. Abst., *39*, #15229	A-N

Engels, J. F., & Wales, H. G. Spoken versus pictured questions on taboo topics. *Journal of Advertising Research*, 1962, *2*(1), 11-17.

Instrument Format	3g
Bureau of Census, #7111011601	A-M

* England, L. R. Capital punishment and open-end questions. *Public Opinion Quarterly*, 1948, *12*, 412-416.

Open-Ended Items, Close-Ended Items, Response Alternatives	2
ORA	R-H

Engvik, H., Kvale, S., & Havik, O. E. Rater reliability in evaluation of essay and oral examinations. Pedagogisk Forskning: Scandinavian Journal of Educational Research, 1970, 4, 195-220.

Erdos, P. L. How to save time and money on the tabulation of surveys. Printers' Ink, 1948 (a), 222(7), 36-37.

* Erdos, P. L. Planning the questionnaire for tabulation. International Journal of Opinion and Attitude Research, 1948 (b), 2(3), 401-408.

Instrument Format, Scoring 8

Bureau of Census, #7110002001 (Rev. from rept.) R-M

Erdos, P. L., & Morgan, A. J. Professional mail surveys. New York: McGraw-Hill, 1970.

ERIC Clearinghouse on Tests, Measurements, and Evaluation. Test bias: A bibliography (ERIC Document Reproduction Service, ED 051 312). Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurements, and Evaluation, 1971.

Response Bias, Investigator Error, Bibliography 12, 16

ORA R-M

Eriksen, C. W., & Hake, H. W. Absolute judgments as a function of stimulus range and number of stimulus and response categories. Journal of Experimental Psychology, 1955, 49, 323-332.

Ericksen, S. C. A skeptical note on the use of attitude scales toward war: I. In 1940, 1941. Journal of Social Psychology, 1942, 16, 229-242.

Evans, F. R., & Reilly, R. A study of speededness as a source of test bias (ERIC Document Reproduction Service, ED 053 207). Princeton, New Jersey: Educational Testing Service, 1971.

* Eysenck, H. J. Response set, authoritarianism and personality questionnaires. British Journal of Social and Clinical Psychology, 1962, 1(1), 20-24.

Response Bias, Attitude Measures 12

Psych. Abst., 37, #1252 A-N

* Eysenck, H. J., & Crown, S. An experimental study in opinion-attitude methodology. International Journal of Opinion and Attitude Research, 1949, 3, 47-86.

Questionnaire Theory and Development, Scaling 2, 14

International Journal of Opinion and Attitude Research, 3, p. 47-86 (Rev. from rept.) R-M

* Eysenck, H. J., & Eysenck, S. B. G. A factorial study of an interview-questionnaire. Journal of Clinical Psychology, 1962, 18(3), 286-290.

Interviews, Personality Measures 1, 7

Psych. Abst., 39, #175 A-H

* Eysenck, S. B. G., & Eysenck, H. J. Acquiescent response set in personality questionnaires. Life Sciences, 1963 (a), No. 2, 144-147.

Response Bias, Personality Measures 2, 12

Psych. Abst., 38, #995 A-M

Eysenck, S. B. G., & Eysenck, H. J. An experimental investigation of "desirability" response set in a personality questionnaire. Life Sciences, 1963 (b), No. 5, 343-355.

Response Bias, Personality Measures 12

Psych. Abst., 38, #2712 A-N

* Faerber, N. N. An experimental study of the relative difficulty of hand-scored versions of the same test. Bulletin of the National Institute of Personnel Research, Johannesburg, 1951, 3(2), 10-19.

Instrument Format, Multiple Choice Items, Open-ended Items 2, 3g, 14

Psych. Abst., 27, #772 A-M

* Falk, G. H., & Bayroff, A. G. Rater and technique contamination in criterion ratings. Journal of Applied Psychology, 1954, 38, 100-102.

Raters, Investigator Error 13

Psych. Abst., 29, #3125 A-H

Falthzik, A. M., & Carroll, S. J. Rate of return for closed versus open ended questions in a mail questionnaire survey industrial organization. Psychological Reports, 1971, 29 (3, Pt. 2), 1121-1122.

* Falthzik, A. M., & Jolson, M. A. Statement polarity in attitude studies. Journal of Marketing Research, 1974, 11, 102-105.

Instrument Format, Question Stem, Response Bias 3g, 12

ORA R-H

Farley, F. H. Global self-ratings, the independence of questionnaire drive and anxiety, and social desirability variance. Acta Psychologica, Amsterdam, 1968, 28(4), 387-397.

Farnsworth, P. R. Shifts in the values of opinion items. Journal of Psychology, 1943, 16, 125-128.

* Farnsworth, P. R. Attitude scale construction and the method of equal-appearing intervals. Journal of Psychology, 1945(a), 20, 245-248.

Attitude Measures, Scaling, Investigator Error, Rating Scales, Clarity 4, 14

ORA R-H

* Farnsworth, P. R. Further data on the obtaining of Thurstone scale values. Journal of Psychology, 1945 (b), 19, 69-73.

Scaling 14

Psych. Abst., 19, #1281 R-H

Fear, R. A., The evaluation interview: Predicting job performance in business and industry. New York: McGraw-Hill, 1958.

Feather, N. T. Test-retest reliability of individual values and value systems. Australian Psychologist, 1971, 6(3), 181-188.

Feder, D. D. Effect of direction and arrangement of items on students' performance in a test. Journal of Educational Research, 1936, 30, 28-45.

Federico, P. Development of psychometric measures of student attitudes toward technical training: Reliability and factorial validity. Brooks AFB, Texas: Air Force Human Resources Laboratory, 1970. Report No. AFHRL-TR-70-37.

Attitude Measures, Military Personnel, Reliability, Scaling, Validity 2

DDC A-M

* Federico, P. Degree of evaluation assertions ascribed to an attitude universe as a function of measurement format (AD 736 788). Psychological Reports, 1971(a), 29, 1315-1324.

Attitude Measures, Scaling 2

DDC A-H

Federico, P. Identifying item validity indices utilizing a multivariate model. Lowry AFB, Colorado: Technical Training Division, Air Force Human Resources Laboratory, 1971 (b). AFHRL-TR-71-16.

N/A N/A

N/A T-M

Fehrer, E. Shifts in scale values of attitude statements as a function of the composition of the scale. Journal of Experimental Psychology, 1952, 44, 179-188.

* Fehrer, E., & Strupp, H. The effect of equating interest test items for prestige value. Journal of Applied Psychology, 1949, 33, 222-230.

Scaling, Interest Measures, Response Bias 8, 12

Psych. Abst., 24, #1923 A-N

Feldman, J. J., Hyman, H., & Hart, C. W. A field study of interviewer effects on the quality of survey data. Public Opinion Quarterly, 1951, 15, 734-761.

Investigator Error, Interviews, Response Bias 12, 13

Psych. Abst., 27, #3451 (Rev. from rept.) R-H

* Feldman, M. J., & Corah, N. L. Social desirability and the forced choice method. Journal of Consulting Psychology, 1960, 24, 480-482.

Response Bias, Forced Choice Items 2, 12

Psych. Abst., 36, #1HF 80F A-M

* Feldman, S. Evaluative ratings of adjective-adjective combinations, predicted from ratings of their components. Dissertation Abstracts International, 1969, 30(2-B), 864.

Adjectives, Response Alternatives, Semantic Differential Items 6, 3b, 3f

Dissertation Abstracts International, 30, 864-B (rev.) A-H

Ferber, R. Order bias in a mail survey. Journal of Marketing, 1952, 17 (2), 171-178.

* Ferber, R. The effect of respondent ignorance on survey results. Journal of the American Statistical Association, 1956, 51, 576-586.

Investigator Error — 9

Psych. Abst., 32, #372 — A-M

* Ferber, R. Item nonresponse in a consumer survey. Public Opinion Quarterly, 1966, 30, 399-415.

Investigator Error — 3c, 9

Potter, Sharpe, Hendee, and Clark, 1972 — R-NA

Ferber, R., & Hauck, M. A framework for dealing with response errors in consumer surveys. Marketing Concept in Action, 1964, 47, 533-540.

* Ferber, R., & Wales, H. G. Detection and correction of interviewer bias. Public Opinion Quarterly, 1952, 16(1), 107-127.

Investigator Error, Interviews — 13

Bureau of the Census, #7111030701 — R-H

Ferber, R., & Wales, H. G. Advertising recall in relation to type of recall. Public Opinion Quarterly, 1958-59, 22(4), 529-536.

Ferber, R., & Wales, H. A basic bibliography on marketing research. Chicago: American Marketing Association, 1963.

Ferguson, H. H. Incentives and an intelligence test. Australian Journal of Psychology, 1935, 15, 39-53.

* Ferguson, L. W. The influence of individual attitudes on construction of an attitude scale. Journal of Social Psychology, 1935 (a), 6, 115-117.

Attitude Measures, Response Bias, Investigator Error, Scaling — 10, 12

ORA — R-M

Ferguson, L. W. *Some problems in the measurement of attitudes*. Unpublished Master's thesis, Stanford University, 1935 (b).

N/A	N/A
N/A	T-M

Ferguson, L. W. *The isolation and measurement of general attitudes*. Paper presented at the Philadelphia meetings of the Eastern Psychological Association, 1939 (a).

N/A	N/A
N/A	T-M

* Ferguson, L. W. The requirements of an adequate attitude scale. *Psychological Bulletin*, 1939 (b), *36*, 665-673.

Attitude Measures, Literature Review, Scaling	16, 14, 15
ORA	R-N

Ferguson, L. W. Comparison of scale values from the method of equal appearing intervals and paired comparisons method. *Journal of General Psychology*, 1940, *23*, 431-435.

Ferguson, L. W. A study of the Likert technique of attitude scale construction. *Journal of Social Psychology*, 1941, *13*, 51-57.

Ferguson, L. W. A revision of the primary social attitude scale. *Journal of Psychology*, 1944, *17*, 229-241.

Ferguson, L. W. Psychology and the Army: Introduction of the Rating Scale. *Heritage of Industrial Psychology*, 1963, *9*, 125-139.

Bibliography	16
Psych. Abst., *38*, #6833	A-M

Ferguson, L. W., Huguenard, T., & Sager, E. B. Interview time, interview set, and interview outcome. *Perceptual and Motor Skills*, 1970, *31*(3), 831-836.

Fernberger, S. W. Instructions and the psychophysical limen. *American Journal of Psychology*, 1931, *43*, 93-100.

Fernberger, S. W., Glass, E., Hoffman, I., & Willig, M. Judgment times of different psychophysical categories. *Journal of Experimental Psychology*, 1934, *17*, 286-293.

Fernberger, S. W., & Irwin, F. W. Time relations for the different categories of judgment in the "absolute method" in psychophysics. *American Journal of Psychology*, 1932, *44*, 505-525.

* Ferris, A. C. A note on stimulating response to questionnaires. *American Sociological Review*, 1951, *16*, 247-249.

Respondents' motivation 11

ORA R-H

Festinger, L. Studies in decision: I. Decision time, relative frequency of judgment, and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, 1943, *32*, 241-306.

Festinger, L. The treatment of qualitative data by "scale analysis." *Psychological Bulletin*, 1947, *44*, 149-161.

Festinger, L., and Katz, D. (Eds.) *Research methods in the behavioral sciences*. New York: Dryden Press, 1953.

* Field, J. B. The effects of praise in a public opinion poll. *Public Opinion Quarterly*, 1955, *19*(1), 85-91.

Interviews, Open-Ended Items, Respondent's Motivation 11

ORA R-H

Fillenbaum, S. The effect of a remote anchor upon judgment with a salient within-series stimulus-object and with a free choice of scale. *American Journal of Psychology*, 1961, *74*, 602-606.

N/A N/A

N/A T-H

Finch, C. W., & Gibson, J. N. *Development of a questionnaire to measure Air Force junior officer's attitudes toward intrinsic aspects of the work itself* (AD 743 405). Wright-Patterson AFB: Air Force Institute of Technology, School of Systems and Logistics, 1971. Report No. SLSR-12-72A.

N/A N/A

N/A T-M

* Findikyan, N. Acquiescence and bipolarity in personality questionnaires. *Dissertation Abstracts*, 1969, *29*(8-B), 3103-3104.

Response Bias, Personality Measures, Rating Scales 2, 12

Dissertation Abstracts, *29*(8-B), p. 3103 (Rev. from rept.) A-N

Fink, H. C. Fictitious groups and the generality of prejudice: An artifact of scales without neutral categories. *Psychological Reports*, 1971, *29*(2), 359-365.

Finn, R. H. Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 1972, *32*(2), 255-265.

Response Alternatives 3a, 3f

Psych. Abst., *49*, #8213 A-H

* Fischer, R. P. Signed versus unsigned personal questionnaires. *Journal of Applied Psychology*, 1946, *30*(3), 220-225.

Anonymous Respondent 11

Potter, Sharpe, Hendee, and Clark, 1972 A-M

Fishbein, M. The relationships between beliefs, attitudes and behavior. In Feldman, S. (Ed.), Cognitive consistency. New York: Academic Press, 1966. Pp. 199-223.

Fishbein, M. Attitude and the prediction of behavior. In Fishbein, J. (Ed.), Readings in attitude theory and measurement. New York: John Wiley, 1967.

Fishbein, M., & Ajzen, I. Attitudes and opinions. Annual Review of Psychology, 1972, 23, 487-544.

Attitude Measures, Literature Review 16

ORA R-N

Fishburn, P. C. Semiorders and risky choices. Journal of Mathematical Psychology, 1968, 5, 358-361.

Fisher, G. A discriminant analysis of reporting errors in health interviews. Applied Statistics, 1962, 11(3), 148-163.

Fisher, H. Interview bias in the recording operation. International Journal of Opinion and Attitude Research, 1950, 4, 391-411.

Fiske, D. W. Consistency of the factorial structures in personality ratings from different sources. (Ph.D. thesis), University of Michigan, 1948.

Fiske, D. W. Subject reactions to inventory format and content. Proceedings of the 77th Annual Convention of the American Psychological Association, 1969, 4, 137-138.

Military Personnel, Instrument Format, Investigator Error 3g

ORA R-H

Fiske, D. W., and Cox, J. A., Jr. The consistency of ratings by peers, Journal of Applied Psychology, 1960, 44, 11-17.

Fiske, D. W., & Pearson, P. H. Theory and techniques of personality measurement. Annual Review of Psychology, 1970, 21, 49-86.

* Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.

Critical Incident Technique	3g, 17
ORA	R-M

* Flanagan, J. C. The development of an index of examinee motivation. Educational and Psychological Measurement, 1955, 15, 144-151.

Response Bias, Military Personnel	11
Educational and Psychological Measurement, 15, p. 150 (Rev.)	R-H

Fleischman, H. L. Consumer acceptance of programmed instruction. St. Louis, Mo.: Washington University, Department of Psychology, 1968. Technical Report No. 15 60-65.

Fleiss, J. L. Analysis of variance methods in assessing errors in interview data. Dissertation Abstracts, 1968, 28(12-B), 5231.

Fleiss, J. L. Estimating the reliability of interview data. Psychometrika, 1970, 35 (2), 143-162.

Fleiss, J. L. Measuring nominal scale agreement among many raters. Psychological Bulletin, 1971, 76(5), 378-382.

N/A	N/A
N/A	A-H

Fletcher, N. C., & Shephard, A. H. Interpretation of data as a function of units of measurement. Canadian Journal of Psychology, 1957, 11, 65-70.

N/A 18

Psych. Abst., 32, #3523 A-N

* Flyer, E. S., & Carp, F. M. The Picture Test: Rationale and one validation of the method. Journal of Applied Psychology, 1962, 46, 226-227.

Response Bias, Rating Scales, Clarity 12, 2, 4

Psych. Abst., 37, #1186 A-M

Fogel, L. J. Biotechnology: Concepts and applications. Englewood Cliffs: Prentice-Hall, 1963.

* Follman, J., Urbanke, R., & Burley, W. Comparison of three matching item formats. Florida Journal of Educational Research, 1971, 13, 39-48.

Matching Items, Instrument Format, Response Alternatives 3b, 3e

ORA R-M

Ford, LeR. H., Jr. A forced-choice, acquiescence-free, social desirability (defensiveness) scale. Journal of Consulting Psychology, 1964, 28(5), 475.

Ford, LeR. H., Jr., & Meisels, M. Social desirability and the semantic differential. Educational and Psychological Measurement, 1965, 25(2), 465-475.

* Ford, N. M. The advance letter in mail surveys. Journal of Marketing Research, 1967, 4, 202.

Respondent's Motivation 11

ORA R-M

* Ford, N. M. Questionnaire appearance and response rates in mail surveys. *Journal of Advertising Research*, 1968, *8*, 43-45.

Respondent's Motivation, Instrument Format 3g, 11

ORA R-M

* Ford, N. M. Consistency of responses in a mail survey. *Journal of Advertising Research*, 1969, *9*(4), 31-33.

Reliability, Interviews, Close-Ended Items 1

ORA R-M

Ford, R. N. Scaling experience by a multiple-response technique: A study of white-Negro contacts. *American Sociological Review*, 1941, *6*, 9-23.

* Ford, R. N. A rapid scoring procedure for scaling attitude questions. *Public Opinion Quarterly*, 1950, *14*, 507-532.

Attitude Measures, Scaling, Data Analysis 15, 14, 8

Psych. Abst., *26*, #2106 R-N

Forthman, J. H. The effects of a zero interval on semantic differential rotated factor loadings. *Journal of Psychology*, 1973, *84*(1), 23-32.

* Foster, R. J. Acquiescent response set as a measure of acquiescence. *Journal of Abnormal and Social Psychology*, 1961, *63*, 155-160.

Response Bias, Personality Measures 12

Psych. Abst., *36*, #4HF55F A-M

* Foster, R. J., & Grigg, A. E. Acquiescent response set as a measure of acquiescence: Further evidence. *Journal of Abnormal and Social Psychology*, 1963, *67*(3), 304-306.

Response Bias 12

Psych. Abst., *38*, #2503 A-N

Fowler, H. M. An application of the Ferguson method of computing item conformity and person conformity. *Journal of Experimental Education*, 1954, *22*, 237-245.

Frank, B. Stability of questionnaire response. *Journal of Abnormal and Social Psychology*, 1935, *30*, 320-324.

Investigator Error, Respondent's Motivation 7, 11

Potter, Sharpe, Hendee and Clark, 1972 A-N

Frankel, L. R. How incentives and subsamples affect the precision of mail surveys. *Journal of Advertising Research*, 1960, *1*(1), 1-5.

Franken, R. B. Formulating questionnaires. *Advertising Fortnightly*, 1924 (a), *2*, 23.

Franken, R. B. Formulating questionnaires. *Advertising Fortnightly*, 1924 (b), *2*, 27-28.

Franzen, R. The construction of a questionnaire. *Market Research*, 1936, *4*(5), 17-19.

Franzen, R. H., & Lazarsfeld, P. F. Mail questionnaire as a research problem. *Journal of Psychology*, 1945, *20*, 293-320.

* Frawley, P. J. *A study of judgment: A factorial analysis of the anchoring effects*. Washington, D. C.: Catholic University of American Press, 1948.

Rating Scales, Data Analysis 3f, 8

ORA R-N

* Frederiksen, N., & Messick, S. *Response set as a measure of personality*. Office of Naval Research Technical Report. Princeton, New Jersey: Educational Testing Service, 1958.

Military Personnel, Response Bias, Investigator Error 12, 7

Psych. Abst., *34*, #4065 A-M

* Freeberg, N. E. Relevance of rater-ratee acquaintance in the validity and reliability of ratings. Journal of Applied Psychology, 1969, 53(6), 518-524.

Investigator Error, Observations, Raters, Validity 7

Journal of Applied Psychology, 53, p. 518 R-NA

Freeman, L. C., & Ataöv, T. Invalidity of indirect and direct measures of attitude toward cheating. Journal of Personality, 1960, 24 (4), 443-447.

Question Stem 3g

Potter, Sharpe, Hendee, and Clark, 1972 A-M

Freidman, M. P., & Fleishman, E. A. A note on the use of a "don't know" alternative in multiple choice tests. Journal of Educational Psychology, 1956, 47, 344-349.

Achievement Measures, Response Alternatives, Reliability 3g

Psych. Abst., 32, #4598 A-N

* French, E. G. A note on the Edwards Preference Schedule for use with basic airmen. Educational and Psychological Measurement, 1958, 18, 109-115.

Investigator Error, Response Bias, Personality Measures, Military Personnel 12, 13, 9

Psych. Abst., 33, #8340 A-M

French, J. W. A note on keying and item-selection for a practical judgment-test. Psychometrika, 1952, 17, 101-6.

French, J. W. The effect of essay test on student motivation. Research Bulletin 56-4. Princeton, New Jersey: Educational Testing Service, 1956.

N/A N/A

N/A T-M

French, J. W. The kinds of items that work in an interest activities index. Research Bulletin 64-36. Princeton, New Jersey: Educational Testing Service, 1964.

N/A	N/A
N/A	T-M

Freyd, M. An appraisal of relative merits of types of rating scales and their use. General Management Series No. 38, New York: American Management Association, 1926.

N/A	N/A
N/A	T-H

Fricke, B. G. Response set as a suppressor variable in the OAIS and MMPI. Journal of Consulting Psychology, 1956, 20, 161-169.

* Fricke, B. G. Subtle and obvious test items and response set. Journal of Consulting Psychology, 1957, 21, 250-252.

Response Bias, Personality Measures, Question Stem	12
Journal of Consulting Psychology, 21, p. 252	R-M

Friesen, E. P. The incomplete sentences technique as a measure of employee attitudes. Personnel Journal, 1952, 5, 329-345.

Frisbie, D. A., & Ebel, R. L. Comparative reliabilities and validities of true-false and multiple choice tests (ERIC Document Reproduction Service, ED 064 388). Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, 1972.

Achievement Measures	18
ERIC Document Reproduction Service, ED 064 388	A-N

* Frisbie, B., & Sudman, S. The use of computers in coding free responses. Public Opinion Quarterly, 1968, 32, 216-232.

Check List, Open-Ended Items, Investigator Error, Scoring 2, 8, 10, 13

OR R-H

Fromme, A. On the use of certain qualitative methods of attitude research. Journal of Social Psychology, 1941, 13, 425-459.

* Fry, E. A readability formula that saves time. Journal of Reading, 1968, 11(7), 575-578

Clarity 3g

ORA A-M

Fry, J. N., & Claxton, J. D. Semantic differential and nonmetric multidimensional scaling descriptions of brand images. Journal of Marketing Research, 1971, 8, 238-240.

Semantic Differential Items, Data Analysis, Card Sorts 2

ORA R-M

Fulkerson, S. C. Individual differences in response validity. Journal of Clinical Psychology, 1959, 15, 169-173.

Fuller, C. H. Effect of anonymity on return rate and response bias in a mail survey (AD-746 478). Washington, D. C.: Naval Personnel Research and Development Laboartories, 1972. Report No. WTR-73-2.

N/A N/A

N/A T-M

Fullerton, G. S., & Cattell, J. McK. On the perception of small differences. Publ. Univ. Penn., Phil. Series, 1892, No. 2.

Furntratt, E. Response tendencies in questionnaires: I. The tendency to agree and the tendency of using the full response scale. *Psychologische Rundschau*, 1969, *20*(1), 1-13.

Furst, E. J. *Constructing evaluation instruments*. New York: David McKay, 1958.

Gadel, M. S. The relationship of item validity shrinkage to curvilinearity of response distributions. *Educational and Psychological Measurement*, 1958, *18*, 145-152.

* Gage, N. L., & Chatterjee, B. B. The psychological meaning of acquiescence set: Further evidence. *Journal of Abnormal and Social Psychology*, 1960, *60*, 280-283.

Response Bias, Investigator Error	12, 3g
Psych. Abst., *34*, #7548	A-M

Gage, N. L., Leavitt, G. S., & Stone, G. C. The psychological meaning of acquiescense set for authoritarianism. *Journal of Abnormal and Social Psychology*, 1957, *55*, 98-103.

* Gaito, J. Forced and free Q sorts. *Psychological Reports*, 1962, *10*(1), 251-254.

Card Sorts, Forced Choice Items, Respondent's Motivation	2, 11
Psych. Abst., *37*, #1187	A-M

Gale, E. J., Jr. *Nonresponse in a mail survey of naval personnel* (AD 733 377). Monterey, California: Naval Postgraduate School, 1971.

N/A	N/A
N/A	T-M

Gales, K., & Kendall, M. G. An inquiry concerning interviewer variability. Journal of the Royal Statistical Society, 1957, Series A, 120 (Part II), 121-147.

Gallup, G. Question wording in public opinion polls: Comments on points raised by Mr. Stegner. Sociometry, 1941, 4, 259-268.

Gallup, G. A guide to public opinion polls. Princeton: Princeton University, 1944.

Gallup, G. Qualitative measurement of public opinion: The quintamensional plan of question design. Princeton, N. J.: American Institute of Public Opinion, 1947 (a).

N/A N/A

N/A T-H

Gallup, G. The quintamensional plan of question design. Public Opinion Quarterly, 1947 (b), 11, 385-393.

Investigator Error 13

Bureau of Census, #7110008301 R-N

Gallup, G. Public opinion surveys. New York, 1950.

Gallup, G., & Rae, S. F. The pulse of democracy. New York: Simon & Schuster, 1940.

Gannon, M. J. Proper use of the questionnaire survey. Business Horizons, 1973, 16, 89-94.

Data Analysis 8

ORA R-N

Garcia-Mata, C. A practical method of smoothing statistical curves by hand. Bull. Pan Amer. Union, 1941, 75, 276-281.

* Gardner, E. F. Comments on selected scaling techniques with a description of a new type of scale. Journal of Clinical Psychology, 1950, 6, 38-43.

Scaling 14

Psych. Abst., 25, #63 A-M

Gardner, P. L. Test length and the standard error of measurement. Journal of Educational Measurement, 1970, 7, 271-274.

Gardner, R. A. Multiple-choice decision-behavior. American Journal of Psychology, 1958, 71, 710-717.

Military Personnel, Multiple Choice Items, Response Alternatives 3a

Psych. Abst., 34, #801 A-H

Garner, W. R. Rating scales, discriminability, and information transmission. Psychological Review, 1960, 67, 343-352.

Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. Psychological Review, 1951, 58, 446-459.

Questionnaire Theory and Development 15

ORA R-N

Garry, R. Individual differences in ability to fake vocational interests. Journal of Applied Psychology, 1953, 37, 33-37.

Garvin, A. D. Non-chance results from a pure-chance test: A study in response position selection set (ERIC Document Reproduction Service, ED 058 284). Paper presented at the Annual Meeting of the American Psychological Association, Washington, D. C., 1971.

Garvin, A. D. *Confidence weighting*. (ED 062 401, MF and HC available from EDRS.)

Garvin, A. D. *Confidence weighting plus Coombs-type response options: A good idea that failed*. (ED 065 551, MF and HC available from EDRS.)

Gatty, R., & Allais, C. *The semantic differential applied to image research*. New Brunswick: Rutgers University (undated).

Gekoski, N. *Psychological testing: Theory, interpretation, and practices*. Springfield, Ill.: Charles C. Thomas, 1969.

Gekoski, N., & Isard, E. S. Note on another use of the sentence-completion technique. *Journal of Applied Psychology*, 1955, 39, 139.

Gelfand, N. I. An investigation of interviewer recall of interviewee responses. *Dissertation Abstracts International*, 1973, 33 (10-B), 5043-5044.

Gentile, J. R., & Seibel, R. A rating scale measure of word-relatedness. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8(2), 252-256.

Data Analysis	15
ORA	R-N

* Georgoff, D. M., Hersker, H. J., & Murdick, R. G. The lost-letter technique -- a scaling experiment. *Public Opinion Quarterly*, 1972, 36(1), 114-119.

Attitude Measures, Investigator Error	10
ORA	R-N

Gerberich, J. B. A study of the consistency of informant responses to questions in a questionnaire. *Journal of Educational Psychology*, 1947, 38, 399-306.

* Gerberich, J. B., & Mason, J. M. Signed versus unsigned questionnaire. *Journal of Education Research*, 1948, *42* (2), 122-126.

Anonymous Respondent	11
Potter, Sharpe, Hendee, and Clark, 1972	A-H

Getzels, J. W., & Walsh, J. J. The method of paired direct and projective questionnaires in the study of attitude structure and socialization. *Psychological Monographs*, 1958, *72* (1) (Whole No. 454).

Ghiselli, E. E. Some further points on public opinion polls. *Journal of Marketing*, 1940, *5*, 115-119.

Question Stem, Clarity, Investigator Error	4, 13
ORA	R-N

Ghiselli, E. E. The problem of question form in the measure of sales by consumer interviews. *The Journal of Marketing*, 1941, *6*, 170-171.

Question Stem, Interviews, Clarity	4, 3g
ORA	R-N

* Ghiselli, E. E. All or none versus graded response questionnaires. *Journal of Applied Psychology*, 1939, *23*, 405-413.

Multiple Choice Items, Response Alternatives	2, 3a
ORA	R-M

Ghiselli, E. E. *The measurement of occupational aptitude*. Berkeley: University of California Press, 1955.

Ghiselli, E. E. *Theory of psychological measurement*. New York: McGraw-Hill, 1964.

* Ghiselli, E. E., & Brown, C. W. *Personnel and industrial psychology*. New York: McGraw-Hill, 1948.

Textbook	17
ORA	R-M

Gibbins, K. Response sets and the semantic differential. *British Journal of Social and Clinical Psychology*, 1968, *7*(4), 253-263.

* Gibson, F. K., & Hawkins, B. W. Interviews versus questionnaires. *American Behavioral Scientist*, 1968, *12*, 9-16.

Interviews	1
ORA	R-M

Gilbert, A. R. Superiority of latency-weighted scores over weighted scores in the assessment of professional interests. *Psychological Reports*, 1970, *26*(1), 93-94.

* Gilbert, T. F. Experiments in morale. *Journal of Social Psychology*, 1956, *43*, 299-308.

Interest Measures, Military Personnel, Response Bias	9, 10, 12
Psych. Abst., *33*, #4801	A-H

Ginter, J. L. An experimental investigation of attitude change and choice of a new brand. *Journal of Marketing Research*, 1974, *XI*, 30-40.

* Githens, W. H. *Influence of questionnaire items on general attitude toward job*. USN PRA Research Memorandum No. 65-13.

Rating Scales, Attitude Measures, Military Personnel, Question Stem	3g
Psych. Abst., *39*, #16 563	A-H

Gividen, G., Nystrom, C. O., & Arsdell, P. M., Jr. Fort Hood semiannual project VOLAR/MVA evaluation report. (Preliminary report.) Office of the Deputy Chief of Staff for Command Programs, III Corps and Fort Hood, Fort Hood, Texas, 1972.

Glaser, R. A methodological analysis of the inconsistency of response to test items. Educational and Psychological Measurement, 1949, 9, 727-740.

Glasser, R. Training research and education. Pittsburgh: University of Pittsburgh Press, 1962.

Gleser, L. J. On bounds for the average correlation between subtest scores in ipsative scoring. Educational and Psychological Measurement, 1972, 32(3), 759-765.

* Glickman, A. S. Studies in career motivation: III. Effects of repeated questionnaire administration on returns and on intended and actual reenlistment (AD 633 598). Washington, D. C.: Bureau of Naval Personnel, Personnel Research Division, 1962. Report No. RS-62-5.

Military Personnel, Respondent's Motivation	11
DDC, #AD 633 598	A-M

* Gloye, E. E. A note on the distinction between social desirability and acquiescent response styles as sources of variance in the MMPI. Journal of Counseling Psychology, 1964, 11(2), 180-184.

Response Bias, Question Stem, Validity, True-False Items	12
Psych. Abst., 39, #5175	A-N

* Goheen, H. W., & Kavruck, S. Selected references on test construction, mental test theory and statistics. Washington, D. C.: U.S. Civil Service Commission, 1950.

Goheen, H. W., and Mosel, J. N. Validity of the employment recommendation questionnaires: II. Comparison with field investigations. *Personnel Psychology*, 1959, *12*, 297-301.

Goldberg, L. R. Objective diagnostic tests and measures. *Annual Review of Psychology*, 1974, *25*, 343-366.

Goldberg, L. R., & Hase, H. D. Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin*, 1967, *67*(4), 231-248.

Golden, G. M., & others. *Item writing rule conformity as related to biographical item response stability*. Paper presented in a session on Testing Materials at the Convention of the Rocky Mountain Psychological Association, Denver, Colorado, May 1971.

Goldfried, M. R., & McKenzie, J. D., Jr. Sex difference in the effect of item style on social desirability and frequency of endorsement. *Journal of Consulting Psychology*, 1962, *26*(2), 126-128.

* Goldsamt, M. R. Effects of scoring method and rating scale length in extreme response style measurement. *Dissertation Abstracts International*, 1972, *32*(10-B), 6030.

Scoring, Rating Scales, Response Alternatives, Response Bias	3a, 8, 9, 12
ORA	R-NA

Goldstein, I. L. The application blank: How honest are the responses? *Journal of Applied Psychology*, 1971, *55* (5), 491-492.

Goldstein, M. J. The social desirability variable in attitude research. *Journal of Social Psychology*, 1960, *52*, 103-108.

Goode, P. V. How to get better results from attitude surveys. Personnel Journal, 1973, 52 (3), 187-192.

Goode, W. J., & Hatt, P. Methods in social research. New York: McGraw-Hill, 1952.

Goodenough, W. A. A technique for scale analysis. Educational and Psychological Measurement, 1944, 4, 179-190.

Goodnow, J. J. Response sequences in a pair of two-choice probability situations. American Journal of Psychology, 1955, 68, 624-630.

Goodstein, L. D., & Heilbrun, A. B. Social desirability response set: Error or predictor variable. Journal of Psychology, 1961, 51, 321-9.

Gorden, R. L. Interviewing: Strategy, techniques, and tactics. Homewood, Ill.: Dorsey, 1969.

* Gordon, L. V. Validities of the forced-choice and questionnaire methods of personality measurement. Journal of Applied Psychology, 1951, 35, 407-412.

Forced Choice Items, Personality Measures, Validity, Closed-Ended Items 2, 8, 14

Psych. Abst., 27, #424 A-M

Gordon, L. V. Validity of scoring methods for bipolar scales. Educational and Psychological Measurement, 1967, 27(4, Pt. 2), 1099-1106.

Gordon, L. V. Are there two extremeness response sets? Educational and Psychological Measurement, 1971, 31(4), 867-873.

* Gordon, M. E. An examination of the relationship between the accuracy of favorability of ratings. *Journal of Applied Psychology*, 1972, *56*(1), 49-53.

Investigator Error	13
Journal of Applied Psychology, *56*, p. 49 (Rev. from rept.)	R-A

Gorfein, D. S. Scaling theory and group influence: A re-examination. *Journal of Social Psychology*, 1964, *62*(2), 303-308.

Scaling, Attitude Measures	3f, 10
Psych. Abst., *39*, #4862	A-M

Gotkin, L. G., & Goldstein, L. S. *Descriptive statistics, a programmed textbook*. New York: John Wiley, 1964.

Gough, H. G. The adjective check list as a personality assessment research technique. *Psychological Reports*, 1960, *6*, 107-122.

Personality Measures	18
Psych. Abst., *34*, #7386	A-N

Gough, H. G., & Heilbrun, A. B. *Adjective Check-list Manual*. Palo Alto: Consulting Psychologists Press, 1965.

Personality Measures	18
ORA	T-H

Gough, H. G., McKee, M. G., & Yandell, R. J. *Adjective checklist analyses of a number of selected psychometric and assessment variables*. Maxwell ARB, Alabama: Officer Education Research Laboratory, 1955. Technical Memorandum OERL-TM-10.

N/A	N/A
N/A	T-H

Gouws, D. J. The problem of testing sequence when administering a battery of tests. Journal of the National Institute of Personnel Research, Johannesburg, 1960, 8, 83-90.

Grannenberg, R. T. The influence of individual attitude and attitude-intelligence upon scale values of attitude items. American Pyschologist, 1955, 10, 330-331. (Abstract)

Grant, D. L. An exploratory study of halo effect in rating. Dissertation Abstracts, 1958, 18, 1096-1099.

* Gray, P. G. The memory factor in social surveys. Journal of the American Statistical Association, 1955, 50, 344-363.

Clarity 4

Psych. Abst., 30, #824 A-H

Green, B. F. Attitude measurement. In Lindzey, G. (Ed.), Handbook of social psychology. Cambridge, Mass.: Addison-Wiley, 1954.

Green, B. F. The sensitivity of Guttman weights. (ED 064 323, MF and HC available from EDRS.)

Green, P. E., & Carmone, F. J. Multidimensional scaling and related techniques in marketing analysis. Boston: Allyn and Bacon, 1970.

Green, P. E., Maheshwari, A., & Rao, V. R. Self-concept and brand preference: An empirical application of multidimensional scaling. Journal of Marketing Research, 1969, 11, 343-360.

* Green, P. E., & Rao, V. A. Rating scales and information recovery - How many scales and response categories to use? Journal of Marketing, 1970, 34, 33-39.

Rating Scales, Response Alternatives 3a

ORA R-M

Green, P., & Tull, D. Research for Marketing Decisions. New York: Prentice-Hall, 1970.

Green, P. E., Wind, Y., & Jain, A. K. Analyzing free-response data in marketing research. Journal of Marketing Research, 1973, 9, 45-52.

* Green, R. F. Does a selection situation induce testees to bias their answers on interest and temperament measures? Educational and Psychological 1951, 11, 503-515.

Preference Measures, Response Bias, Investigator Error 12, 7

ORA R-M

Green, R. F., & Goldfried, M. R. On the bipolarity of semantic space. Psychological Monographs: General and Applied, 1965, 79 (6, Whole No. 599).

Semantic Differential Items 18

Psych. Abst., 39, #9441 A-N

Green, R. T., & Stacey, B. G. The response style myth: An empirical study involving the T-scale. Acta Psychologica, Amsterdam, 1966, 25(4), 365-372.

Greenberg, A. L. Respondent ego-involvement in large scale surveys. Journal of Marketing, 1956, 20, 390-393.

Interviews, Respondent's Motivation, Projective Items, Open-Ended Items, X-O Test Items 1, 2, 11

ORA R-M

Greenberg, A. Validity of a brand awareness question. Journal of Marketing, 1958, 23, 182-184.

* Greenberg, A. Pictorial stereotypes in a projective test. Journal of Marketing, 1959, 23, 72-74.

Projective Items	15
ORA	R-N

* Greenberg, A. Paired Comparisons vs. monadic tests. Journal of Advertising Research, 1963, 3(4), 44-47.

Paired Comparison Items, Rating Scales, Preference Measures	2, 3c
ORA	R-H

Greenberg, M. G. A modification of Thurstone's Law of Comparative Judgment to accommodate a judgment category of 'equal' or 'no difference'. Psychological Bulletin, 1964, 64, 108-12.

Response Alternatives	3g
Psychological Bulletin, 64, p. 108	R-NA

* Greenwald, H. J., & O'Connell, S. M. Comparison of dichotomous and likert formats. Psychological Reports, 1970, 27(2), 481-482.

True-False Items, Rating Scales, Reliability	2
ORA	R-H

Greenwood, J. A. A preferential matching problem. Psychometrika, 1943, 8, 185-191.

Gregson, R. A. Representation of taste mixture cross-modal matching in a Minkowski r-metric. Australian Journal of Psychology, 1965, 17(3), 195-204.

Paired Comparison Items, Clarity, Instrument Format, Matching Items, Data Analysis	4, 2, 3g, 8
Psych. Abst., 40, #4734	A-M

Grice, H. H. The construction and validation of a generalized scale to measure attitude toward defined groups. Bulletin Purdue University, 1934, 25, 37-46.

Gridgeman, N. T. Significance and adjustment in paired comparisons. Biometrics, 1963, 19(2), 213-228.

Gritten, F., & Johnson, D. M. Individual differences in judging multiple-choice questions. Journal of Educational Psychology, 1941, 32, 423-430.

* Gross, E. J. The effect of question sequence on measures of buying interest. Journal of Advertising Research, 1964, 4, 41.

Instrument Format, Preference Measures, Investigator Error	3c
ORA	R-H

* Guber, J. F., & Gerberich, J. B. A note on consistency in questionnaire responses. American Sociological Review, 1946, 11(1), 13-15.

Investigator Error	13
Bureau of Census, #7110000601 (Rev.)	A-H

Guest, L. A study of interviewer competence. International Journal of Opinion and Attitude Research, 1947, 1(4), 17-30.

Guest, L. A comparison of two-choice and four-choice questions. Journal of Advertising Research, 1962 (a), 2, 32-34.

Instrument Format, Multiple Choice Items, Response Alternatives	3a, 2
ORA	R-H

291

Guest, L. Consumer analysis. Annual Review of Psychology, 1962 (b), 13, 315-344.

Guest, L., & Nuckols, R. A laboratory experiment in recording in public opinion interviewing. International Journal of Opinion and Attitude Research, 1950, 4, 336-352.

* Guilford, J. P. The method of paired comparisons as a psychometric method. Psychological Review, 1928, 35, 494-506.

Questionnaire Theory and Development, Scaling 14

Psychological Review, 35, p. 506 R-N

Guilford, J. P. The difficulty of a test and its factor composition. Psychometrika, 1941, 6, 67-77.

* Guilford, J. P. Psychometric methods. (2nd ed.) New York: McGraw-Hill, 1954.

Textbook 17

ORA R-H

Guilford, J. P. Personality. New York: McGraw-Hill, 1959.

Guilford, J. P. Fundamental statistics in psychology and education. (4th ed.) New York: McGraw-Hill, 1965.

Guilford, J. P., & Jorgensen, A. P. Some constant errors in ratings. Journal of Experimental Psychology, 1938, 22, 43-57.

Scaling 14

Psych. Abst., 12, #2289 (Rev. from rept.) R-M

Guilford, J. P., & Lacey, J. I. Printed classification tests. Washington, D. C.: Government Printing Office, 1947. Research Report No. 5, AAF Psychology Program.

Guinn, N. Development of non-cognitive measures for use in officer selection (DF231630). Lackland AFB, Texas: Air Force Human Resources Laboratories, 1973.

N/A	N/A
N/A	R-NA

Gullahorn, J. E., & Gullahorn, J. T. An investigation of the effects of three factors on response to mail questionnaire. Public Opinion Quarterly. 1963, 27, 294-296.

Gulliksen, H. Paired comparisons and the logic of measurement. Psychological Review, 1946, 53, 199-213.

Questionnaire Theory and Development	18
ORA	R-N

Gulliksen, H. The reliability of a speeded test. American Psychologist, 1949, 4, 243.

Gulliksen, H. Theory of mental tests. New York: John Wiley, 1950.

Gulliksen, H. Measurement of subjective values. Psychometrika, 1956, 21, 229-244.

Questionnaire Theory and Development, Data Analysis	15
Psychometrika, 33, p. 229	R-N

Gulliksen, H. How to make meaning more meaningful. Contemporary Psychology, 1958, 3, 115-119.

Semantic Differential Items, Data Analysis, Scaling 3a, 8

ORA R-N

Gulliksen, H. Attitudes of different groups toward work, aims, goals, and activities. Princeton, N. J.: Princeton University, Dept. of Psychology. Contract N00014-67-A-0151-0006.

* Gulliksen, H., & Messick, S. (Eds.) Psychological scaling: Theory and applications. New York: John Wiley and Sons, 1969.

Scaling 14

ORA R-H

Gulliksen, H., & Tucker, L. R. A general procedure for obtaining paired comparisons from multiple rank orders. Psychometrika, 1961, 26, 197-183.

* Gustav, A. Students' preferences for test format in relation to their test scores. Journal of Psychology, 1964, 57(1), 159-164.

True-False Items, Multiple Choice Items, Open-Ended Items 2, 9, 10

Psych. Abst., 39, #2890 A-N

Guttman, L. A basis for scaling qualitative data. American Sociological Review, 1944, 9, 139-150.

Questionnaire Theory and Development, Scaling 14

ORA R-N

* Guttman, L. The Cornell technique for scale and intensity analysis. Educational and Psychological Measurement, 1947 (a), 7, 247-280.

Scaling, Scoring 14, 8

ORA R-M

Guttman, L. Suggestions for further research in scale and intensity analysis of attitudes and opinions. International Journal of Opinion and Attitude Research, 1947 (b), 1(1), 30-35.

Guttman, L. On Festinger's evaluation of scale analysis. Psychological Bulletin, 1949, 44, 451-465.

Literature Review, Scaling, Questionnaire Theory and Development 15, 16

ORA R-N

Guttman, L. Measuring the true state of opinion. Ferber, R., & Wales, H. (Eds.), Motivation and Behavior. Homewood, Ill.: Richard D. Irwin, 1958.

Guttman, L. The basis for scalogram analysis. In Stouffer S. A. (Ed.), Measurement and prediction. Princeton, N. J.: Princeton University Press, 1950.

Guttman, L., & Suchman, E. A. Intensity and a zero point for attitude analysis, American Sociological Review, 1947, 12, 57-67.

Attitude Measures, Data Analysis, Questionnaire Theory and Development, Response Bias, Military Personnel 12, 2, 8

ORA R-M

Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. Journal of Psychology, 1949, 27, 453-463.

Adjectives, Paired Comparison Items 6

ORA R-M

* Haire, M. Projective techniques in marketing research. Journal of Marketing, 1950, 14, 649-656.

Projective Items 2

ORA R-M

Hake, H. W., & Garner, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. Journal of Experimental Psychology, 1951, 42, 358-366.

Response Alternatives	3a, 8
Journal of Experimental Psychology, 42, pp. 365-366 (Rev.)	R-M

Hall, E. Item writing in empirical studies. Educ. Res., 1969, 11(3), 223-225.

Hall, R. F. An application of unfolding theory to the measurement of attitudes. Educational and Psychological Measurement, 1970, 30(3), 621-37.

Haller, T. P. Let's not bury paired comparisons. Journal of Advertising Research, 1966, 6(3), 29-30.

Paired Comparison Items	2
ORA	R-N

Halpern, G. Item arrangement and bias in an interest test. Research Bulletin 67-39. Princeton, New Jersey: Educational Testing Service, 1967.

N/A	N/A
N/A	T-H

Halpern, R. S. Some observations about attitudes, attitude measurement, and behavior. In adler, L., & Crespi, J. (Eds.), Attitude Research on the Racks. Chicago: American Marketing Association, 1968.

Hambleton, R. K. The effects of item order and anxiety on test performance and stress (ERIC Document Reproduction Service, Ed 179 960). Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1968.

Achievement Measures	18
ERIC Document Reproduction Service, ED 017 960	A-M

* Hamel, L., & Reif, H. G. Should attitude questionnaires be signed? Personnel Psychology, 1952, 5(2), 78-91.

Anonymous Respondent, Investigator Error   11, 9

Psych. Abst., 27, #4646   R-H

Hamid, P. N. Word meanings and self-descriptions. Journal of Social Psychology, 1969, 79, 51-54.

Hamilton, C. H. Bias and error in multiple-choice tests. Psychometrika, 1950, 15, 151-168.

Data Analysis, Response Alternatives   8, 3a

Psych. Abst., 25, #693   A-N

Hammock, J. C. Anxiety scales for use in army training research (AD480314) Human Resources Research Organization, Training Methods Division, Staff Memorandum, 1954.

Hammond, K. R. Measuring attitudes by error-choice: An indirect method. Journal of Abnormal and Social Psychology, 1948, 9, 38.

Hancock, J. W. An experimental study of limiting response on attitude scales. In Remmers, H. H. (Ed.), Further Studies in Attitudes, Series III. Studies in Higher Education, XXXIV. Lafayette, Ind.: Purdue University, 1938. Pp. 142-148.

N/A   N/A

N/A   T-H

Hancock, J. W. An experimental study of four methods of measuring unit cost of obtaining attitudes toward a retail store. Journal of Applied Psychology, 1940, 24, 213-230.

Hand, J. Comment on "Acquiescence, social desirability and inhibition reflected by 'response set' scales." Psychological Reports, 1963, 13(3), 662.

Hand, J. Measurement of response sets. Psychological Reports, 1964, 14(3), 907-913.

Handy, U., & Lentz, T. F. Item value and test reliability. Journal of Educational Psychology, 1934, 25, 703-708.

Reliability	5, 14
ORA	R-N

Handyside, J. A. A general introduction to attitude scaling. London: Market Research Society, 1960.

Hanley, C. Deriving a measure of test-taking defensiveness. Journal of Consulting Psychology, 1957, 21, 391-397.

Hanley, C. Responses to the wording of personality test items. Journal of Consulting Psychology, 1959, 23, 261-265.

Hanley, C. Personality item difficulty and acquiescence. Journal of Applied Psychology, 1965, 49(3), 205-208.

Response Bias, Investigator Error, Personality Measures	12, 8
Psych. Abst., 39, #12286	A-H

Hansen, M. H., Horowitz, W. N., & Madow, W. G. Sample survey methods and Theory. Vol. I. New York: John Wiley, 1960.

Hansen, M. H., & Steinberg, J. Control of error in surveys. Biometrics, 1956, 12, 462-474.

* Hanson, R. H., & Marks, E. S. Influence of the interviewer on the accuracy of survey results. Journal of the American Statistical Association, 1958, 53, 635-655.

Clarity, Interviews, Investigator Error 4,13

Psych. Abst., 33, #10.132 A-M

Hardin, E., & Hershey, G. L. Accuracy of employee reports on changes in pay. Journal of Applied Psychology, 1960, 44, 269-275.

Harding, F. D., & Madden, J. M. Analysis of some aspects of the Air Force position evaluation system. USAF WADD Personnel Lab. Technical Note No. 60-143, 1960.

N/A N/A

N/A T-M

Harding, F. D., Madden, J. M., & Colson, K. Analysis of a job evaluation system. Journal of Applied Psychology, 1960, 44, 354-357.

Harding, L. W. A value-type generalizations test. Journal of Social Psychology, 1944 (a), 19, 53-75.

Harding, L. W. A value-type problemmaire. Journal of Social Psychology, 1944 (b), 19, 115-144.

Hardt, F. A. Comparability and accuracy of interview and questionnaire. Catalogue of Selected Documents in Psychology, 1971, 26 (1).

Hardt, R. H., & Bodine, G. E. Development of self-report-instruments in delinquency research: A conference report. Syracuse, New York: Syracuse University, 1965.

Hare, P. A. Interview responses: Personality or conformity? *Public Opinion Quarterly*, 1960, *24*(4), 679-685.

Attitude Measures, Interviews, Response Bias, Validity, Investigator Error — 7, 10, 12, 9

ORA — R-M

Harmon, L. W. Sexual bias in interest measurement. *Measurement and Evaluation in Guidance*, 1973, *5*(4), 496-501.

Harms, T. Evaluating settings for learning. *Young Children*, 1970, *25*(5), 304-308.

Harris, F. J., Howell, M. A., & Newman, S. H. Forced choice tetrads-effect of scoring procedure and key length on validity and reliability. *Educational and Psychological Measurement*, 1956, *16*(4), 454-464.

Harris, M., & Connelly, G. M. Introducing a symposium on interviewing problems. *International Journal of Opinion and Attitude Research*, 1948, *2*(1), 70-84.

Harris, R. J. Deterministic nature of probabilistic choices among identifiable stimuli. *Journal of Experimental Psychology*, 1969, *79*(3, Pt. 1), 552-560.

Harrison, R. L. Workers' perceptions and job success: Use of forced-choice questionnaires. *Personnel Psychology*, 1959, *12*, 619-25.

N/A — N/A

ORA — R-N

Hart, C. W. Bias in interviewing in studies of opinions, attitudes, and consumer wants. *Proceedings of the American Philosophical Society*, 1948, *92*(5), 399-404.

Response Bias, Investigator Error, Interviews — 12, 13

ORA — 300 — R-N

* Hart, G. L., Faust, R. A., Jr., Rowland, G. E., & Lucier, R. O. Attitudes of troops in the tropics. Volume II. Methodological implications. Haddonfield, New Jersey: Rowland, 1964. R&C Report No. 64-19.

Attitude Measures, Projective Items, Check List, Military Personnel, Respondent's Motivation, Rating Scales, Semantic Differential Items, Open-Ended Items 2, 11, 9, 3g

ORA R-H

Hartson, L. D. Influence of level of motivation on the validity of intelligence tests. Educational and Psychological Measurement, 1945, 5, 273-283.

Hasler, K. R. Importance of descriptive validity in psychological measurement. Personnel Journal, 1972, 51 (1), 12-16.

Hastorf, A. H., & Piper, G. W. A note on the effect of explicit instructions on prestige suggestion. Journal of Social Psychology, 1941, 33, 289-293.

Attitude Measures, Response Bias, Investigator Error 12, 13, 7

Psych. Abst., 26, #6069 A-M

Hauck, M. Is survey postcard verification effective? Public Opinion Quarterly, 1969, 33(1), 117-120.

Hawkins, L., & Coble, J. A. The problem of response error in interviews-Chapter 3. In Lansing, J. B., Withy, S. B., Wolfe, A. C., et al., Working papers on survey research in poverty areas. Ann Arbor, Mich.: University of Michigan, Institute for Social Research, Survey Research Center, 1971.

Hay, N. Establishing factor scales. Personnel, 1946, 23(2).

Hayes, D. P. Item order and Guttman scales. *The American Journal of Sociology*, 1964, *70*(1), 51-58.

Hays, W. L. *Quantification in psychology*. Belmont, Calif.: Brooks/Cole, 1967.

Hays, W. L., & Bennett, J. F. Multidimensional unfolding: Determining configuration from complete rank order preference data. *Psychometrika*, 1961, *26*, 221-238.

Hazlett, C. B. MEDSIRCH: Multiple choice test items. *Educational Technology*, 1973, *13*(3), 23-24.

Heaps, R. A. Use of the semantic differential technique in research: Some precautions. *Journal of Psychology*, 1972, *80*, 121-25.

* Heilbrun, A. B., Jr. Relationships between the Adjective Check-List, Personal Preference Schedule and desirability factors under varying defensiveness conditions. *Journal of Clinical Psychology*, 1958, *14*, 283-287.

Adjectives, Personality Measures, Response Bias 6, 12

Psych. Abst., *33*, #8351 A-M

Heilbrun, A. B., Jr. Validation of a need scaling technique for the Adjective Check List. *Journal of Consulting Psychology*, 1959, *23*, 347-351.

* Heilburn, A. B., Jr. Evidence regarding the equivalence of ipsative and normative personality scales. *Journal of Consulting Psychology*, 1963, *27*, 152-6.

Scaling, Personality Measure, Check List, Forced Choice Items 14

Psych. Abst., *37*, #7982 A-M

Heilbrun, A. B., & Goodstein, L. D. Relationships between personal and social desirability sets and performance on the Edwards Personal Preference Schedule. *Journal of Applied Psychology*, 1959, *43*, 302-305.

Response Bias 12

Psych. Abst., *34*, #5719 A-N

Heise, D. R. Some methodological issues in semantic differential research. *Psychological Bulletin*, 1969, *72*, 406-422.

Heller, K., Davis, J. D., & Myers, R. A. The effects of interviewer style in a standardized interview. *Journal of Consulting Psychology*, 1966, *30*(6), 501-508.

Helme, W. H. *Validation of interest scales for construction and mechanical jobs*. Washington, D. C.: Army Personnel Research Office, 1967.

Interest Measures 18

NTIS A-N

Helme, W. H., Kotula, L. J., & Fitch, D. J. *Preliminary evaluation of measures to predict Army reenlistment*. USA TAG, R&D Command, Human Factors Research Branch, 1960. Technical Research Note No. 110.

Military Personnel, Scoring 14

Psych. Abst., *36*, #2LD14H A-N

Helson, H. An experimental investigation of the effectiveness of the "big lie" in shifting attitudes. *Journal of Social Psychology*, 1958, *43*, 51-60.

Helson, H., & Kozaki, A. Anchor effects using numerical estimates of simple dot patterns. *Perceptual Psychophysiology*, 1968, *4*, 163-164.

Investigator Error, Clarity 3f, 4

Perception and Psychophysics, *4*, p.163 (Rev. from rept.) R-H

Helson, H., Michels, W. C., & Sturgeon, A. The use of comparative rating scales for the evaluation of psychophysical data. American Journal of Psychology, 1954, 67, 321-326.

Hembree, H. W. Dimensionality of soldier acceptance: An approach to critical research. College Park, Maryland: University of Maryland. Technical Report No. 10, QM Contract DA 44-109-QM-129.

N/A N/A

N/A T-M

Hemkee, H. W., & McDermott, W. G. An experimental investigation of synthetic ruffs. Quartermaster General, R&D Division, Environmental Protection Branch, 1952. Report No. 196-DA.

N/A N/A

N/A T-M

Hendrick, C., et al. Effectiveness of ingratiation tactics in a cover letter on mail questionnaire response. Psychonomic Science, 1972, 26 (6), 349-351.

Hendrickson, G. F. The effect of differential option weighting on multiple-choice objective tests. Journal of Educational Measurement, 1971, 8, 291-296.

Heneman, H. G., Jr., & Paterson, D. G. Refusal rates and interviewer quality. International Journal of Opinion and Attitude Research, 1949, 3(3), 392-398.

Henmon, V. A. C. The relation of the time of a judgment to its accuracy. Psychological Review, 1911, 18, 186-201.

Henry, A. F., & Borgatta, E. F. A consideration of some problems of content identification in scaling. Public Opinion Quarterly, 1956, 20, 457-469.

304

Henry, H. Belson's studies in readership. *Journal of Advertising Research*, *2*(2), 9-14.

N/A 18

ORA A-NA

* Heron, A. The effects of real-life motivation on questionnaire response. *Journal of Applied Psychology*, 1956, *40* (2), 65-68.

Card Sorts, Personality Measures, Respondent's Motivation, Investigator Error 3g, 7, 10, 11

Psych. Abst., *31*, #6731 (Rev. from rept.) R-M

* Hess, R. D., & Hink, D. L. A comparison of forced vs. free Q-sort procedure. *Journal of Educational Research*, 1959, *53*, 83-90.

Card Sorts 2

Psych. Abst., *35*, #786 (Rev. from rept.) R-H

Hevner, K. An empirical study of three psychophysical methods. *Journal of General Psychology*, 1930, *4*, 191-212.

Hevner, K. A method of correcting for guessing in true-false tests and empirical evidence in support of it. *Journal of Social Psychology*, 1932, *3*, 359-362.

* Hicks, L. E. Some procedures of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 1970, *74*, 167-184.

Forced Choice Items, Investigator Error, Literature Review, Scoring 2, 13, 8

Psychological Bulletin, *74*, p. 167 (Rev. from rept.) R-H

Highland, R. W., & Berkshire, J. R. A methodological study of forced-choice performance ratings. Bulletin No. 51-59. USAF Human Resources Research Center, 1951.

* Hildum, D. C., & Brown, R. W. Verbal reinforcement and interviewer bias. Journal of Abnormal and Social Psychology, 1956, 53(1), 108-111.

Interviews, Investigator Error, Respondent's Motivation	13, 11
Psych. Abst., 32, #1480	A-M

* Hill, R. J. A note on inconsistence in paired comparison judgments. American Sociological Review, 1953, 18, 564-566.

Paired Comparison Items, Response Bias	8, 10, 12
Psych. Abst., 28, #5185 (Rev. from rept.)	R-H

* Hillmer, M. L., Jr. Social desirability in a two-choice personality scale. (Doctoral dissertation, University of Washington) Seattle, Washington, June 1958.

Response Bias, Personality Measures, Rating Scales, True-False Items	12, 2
ORA	A-NA

Hills, S. L. Research note increasing the response rate for structured interviews in community research. American Behavioral Scientist, 1968, 11(3), 47-48.

Hilton, D. W. Response sets as they relate to item direction in an adjective rating instrument. Dissertation Abstracts International, 1970, 31(5-A), 2182.

Himelstein, P., & Blaskovics, T. L. Prediction of an intermediate criterion of combat effectiveness with a biographical inventory. Journal of Applied Psychology, 1960, 44, 166-168.

* Hinckley, E. D. The influence of individual opinion on construction of an attitude scale. *Journal of Social Psychology*, 1932 (a), *3*, 283-296.

Attitude Measures, Card Sorts, Investigator Error, Scaling	12, 10
ORA	R-M

Hinckley, E. D. *The influence of individual opinion on construction of an attitude scale*. Chicago: University of Chicago Library, 1932 (b).

Hinckley, E. D. A follow-up study on the influence of individual opinion on the construction of an attitude scale. *Journal of Abnormal and Social Psychology*, 1963, *67*, 290-292.

* Hinrichs, J. R., & Gatewood, R. D. Differences in opinion-survey response patterns as a function of different methods of survey administration. *Journal of Applied Psychology*, 1967, *51* (6), 497-502.

Instrument Format, Investigator Error	3g, 7
Journal of Applied Psychology, *51*, p. 497 (Rev. from rept.)	R-H

Hochstim, J. R. A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association*, 1967, *62*(319), 976-989.

Hofstaetter, P. R. Importance and actuality. *International Journal of Opinion and Attitude Research*, 1951, *5*, 31-52.

Attitude Measures, Multiple Choice Items, Rating Scales, Data Analysis, Question Stem	3a, 10
Psych. Abst., *26*, #4738 (Rev. from rept.)	R-N

Hofstaetter, P. The actuality measure in the study of public opinion. *Journal of Applied Psychology*, 1953, *37*, 281-287.

307

Hofstee, W. K. Set scores and forced-choice tests. Research Bulletin 64-38. Princeton, New Jersey: Educational Testing Service, 1964.

N/A	N/A
N/A	T-H

Hofstee, W. K. Secular trends in an adjective checklist. Educational and Psychological Measurement, 1966, 26, 363-367.

Instrument Format, Paired Comparison Items, Response Alternatives	3c
Educational and Psychological Measurement, 26, p. 367 (Rev. from rept.)	R-H

Hofstee, W. K. Comparative vs. absolute judgments of trait desirability. Educational and Psychological Measurement, 1970, 30, 639-646.

Hohle, R. H. Some empirical applications and evaluations of two models for discriminability scales: Pair comparisons and successive intervals. Dissertation Abstracts, 1962, 22(7), 2476-2477.

Holdaway, E. A. Different response categories and questionnaire response patterns. The Journal of Experimental Education, 1971, 40, 57-60.

Instrument Format, Response Alternatives	3g, 3f
Journal of Experimental Education, 40(2), p. 57 (Rev. from rept.)	R-H

Holdrege, F. E., Jr. A combination of forced choice and check list rating scales for the evaluation of instrument flying proficiency. Dissertation Abstracts, March 1960, 20, 3817-3818.

Hollingworth, H. L. Experimental studies in judgment. Archives of Psychology, New York 1913, No. 29.

* Hollis, J. R. Evaluation of a vehicle rating scale. U. S. Army Ordnance Human Engineering Laboratory Technical Memo No. 8, 1954.

Military Personnel, Investigator Error 9

Psych. Abst., 31, #1903 A-H

Holzinger, K. J. The reliability of a single test item. Journal of Educational Psychology, 1932, 23, 211-417.

* Horn, J. L., & Cattell, R. B. Vehicles, ipsatization, and the multiple-method measurement of motivation. Canadian Journal of Psychology, 1965, 19 (4), 265-279.

Personality Measures, Response Bias, Reliability 2, 12

Psych. Abst., 40, #2903 (Rev.) A-M

Horowitz, J. L., et al. Repeated measures effects in racial attitude measurement. College Park, Md.: Maryland University, Cultural Study Center, 1972. Report No. RR-8-72.

Horst, A. P. A method for determining the absolute affective value of a series of stimulus situations. Journal of Educational Psychology, 1932, 23, 418-440.

Horst, P. The chance element in the multiple-choice test item. Journal of General Psychology, 1932 (a), 6, 209-211.

Horst, P. The difficulty of multiple-choice test item alternatives. Journal of Experimental Psychology, 1932 (b), 15, 469-472.

Horst, P. Obtaining a composite measure of a number of different measures of the same attribute. Pyschometrika, 1936, 1, 53-60.

* Horst, P., & Wright, C. E. The comparative reliability of two techniques of personality appraisal. Journal of Clinical Psychology, 1959, 15(4), 388-391.

Paired Comparison Items, Rating Scales, Reliability, Personality Measures, Forced Choice Items, Response Bias   2, 5, 12

ORA   R-H

Hovland, C. I., & Sherif, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. Journal of Abnormal and Social Psychology, 1952, 47, 822-832.

* Howe, E. S. Further comparison of two short-item derivatives of the Taylor Manifest Anxiety Scale. Psychological Reports, 1960, 6, 21-22.

Forced-Choice Items, Response Bias   2, 12

Psych. Abst., 34, #7848   A-H

* Hubbard, A. W. Phrasing questions. Journal of Marketing, 1950, 15, 48-56.

Instrument Format, Question Stem, Reliability   3g, 4, 17

Potter, Sharpe, Hendee, and Clark, 1972 (Rev. from rept.)   R-M

Hubbard, F. W. Questionnaires. Review of Educational Research, 1939, 9, 502-507.

Hubbard, F. W. Questionnaires, interviews, personality schedules. Review of Educational Research, 1942, 12 (5), 534-541.

Literature Review   16

Potter, Sharpe, Hendee & Clark, 1972   A-H

* Hughes, G. D. A new tool for sales managers. Journal of Marketing Research, 1964, 1, 32-38.

Check List, Semantic Differential Items   17

ORA   R-N

* Hughes, G. D. Selecting scales to measure attitude change. Journal of Marketing Research, 1967, 4, 85.

Attitude Measures, Check List, Rating Scales, Respondent's Motivation, Semantic Differential Items, Reliability, Validity	2,11
ORA	R-H

* Hughes, G. D. Some confounding effects of forced-choice scales. Journal of Marketing Research, 1969, 6, 223-226.

Forced Choice Items, Instrument Format, Literature Review, Response Alternatives, Semantic Differential Items	1, 2, 3a
ORA	R-H

Humm, D. G., & Humm, K. A. Compensations for subjects' response-bias in a measure of temperament. American Psychologist, 1947, 2, 305.

Humm, D. G., & Humm, K. A. Notes on "The validity of personality inventories in military practice," by Ellis and Conrad. Psychological Bulletin, 1949, 46, 303-306.

Hummel, R. C. Interviewee responsiveness as a function of interviewer method. Dissertation Abstracts, 1959, 19, 1846.

Humphrey, R. L. Troop-community relations research in Korea: Educational materials (AD 865 725L). Silver Spring, Maryland: American Institutes for Research, 1969. Report No. AIR-R69-15, AIR-E35-S/69-EM.

N/A	18
ORA	R-NA

* Hunt, W. A. Anchoring effects in judgment. American Journal of Psychology, 1941, 54, 395-403.

Rating Scales	3f
American Journal of Psychology, 54, p. 403	R-H

* Hunt, W. A., & Volkmann, J. The anchoring of an affective scale. American Journal of Psychology, 1937, 49, 88-92.

Rating Scales 3f

ORA R-M

Hurd, A. W. Comparisons of short answer and multiple-choice tests covering identical subject content. Journal of Educational Research, 1932, 26, 28-30.

Husek, T. R. Acquiescence as a factor in test-taking behavior and as a personality characteristic. Dissertation Abstracts, 1959, 19, 2650-2651.

Husek, T. R. Acquiescence as a response set and as a personality characteristic. Educational and Psychological Measurement, 1961, 21(2), 295-307.

Hutchinson, B. Some problems of measuring the intensiveness of opinion and attitudes. International Journal of Opinion and Attitude Research, 1949, 3(1), 123-131.

Hyman, H. Do they tell the truth? Public Opinion Quarterly, 1944, 8, 557-559.

Hyman, H. Inconsistencies as a problem in attitude measurement. Journal of Social Issues, 1949(a), 5(3), 32-42.

Attitude Measures, Investigator Error 13

Psych. Abst., 25, #1018 A-N

Hyman, H. Isolation, measurement, and control of interview effect. Social Sciences Research Council, 1949 (b), 3, 15-17.

Hyman, H. H. Problems in the collection of opinion research data. American Journal of Sociology, 1950, 55, 362-370.

Hyman, H. H. Survey design and analysis: Principles, cases, and procedures. Glencoe, Ill.: Free Press, 1955.

Hyman, H. H., Cobb, W. J., Feldman, J. J., Hart, C. W., & Stember, C. H. (Eds.) Interviewing in social research. Chicago: University of Chicago Press, 1954.

Industrial Relation Center, University of Minnesota. Validity of work histories obtained by interview. The Minnesota Studies in Vocational Rehabilitation, 1961.

The interview in social research. American Journal of Sociology, 1956, 62 (2), 137-194.

* Isard, E. S. The relationship between item ambiguity and discriminating power in a forced-choice scale. Journal of Applied Psychology, 1956, 40, 266-268.

Clarity, Response Bias, Forced Choice Items, Rating Scales 4, 12

ORA R-M

Issac, S., and Michael, W. B. Handbook in research and evaluation. San Diego, Calif.: Robert R. Knapp, 1971.

Ito, R. An analysis of response errors - A case study. Journal of Business, 1963, 36(4), 440-447.

Ivens, S. H. Nonparametric item evaluation index. Educational and Psychological Measurement, 1971, 31(4), 831-842.

* Izard, C. E., & Rosenberg, N. Effectiveness of a forced-choice leadership test under varied experimental conditions. Educational and Psychological Measurement, 1958, 18, 57-62.

Military Personnel, Forced Choice Items, Reliability, Validity, Response Bias 2, 12

Psych. Abst., 33, #7293 A-M

Jackson, C.A., Woods, A. R., Stockston, T. W., & Schoof, M. W. OSD Reserve Component Study, Ft. Hood, Texas. Test 6, Phase I. Criteria Validation Report. Washington, D. C.: Department of the Army, 1972. 4 vols.

Military Personnel, Questionnaire Theory and Development — 17

NTIS — A-M

Jackson, D. N. Acquiescence response styles: Problems of identification and control. In Berg, I. A. (Ed.), Response set in personality assessment. Chicago: Aldine, 1967. Pp. 71-114.

Jackson, D. N., Guthrie, G. M. Multitrait-multimethod evaluation of the Personality Research Form. Proceedings of the American Psychological Association, 1968, 177-178.

* Jackson, D. N., & Messick, S. J. A note on "ethnocentrism" and acquiescent response sets. Journal of Abnormal and Social Psychology, 1957, 54, 132-134.

Response Bias, Attitude Measures, Instrument Format — 12, 14

Psych. Abst., 33, #1279 — A-N

Jackson, D. N., & Messick, S. Response styles and the assessment of psychopathology. In Messick, S., & Ross, J. (Eds.), Measurement in personality and cognition. New York: John Wiley, 1962. Pp. 129-155.

Jackson, D. N., & Messick, S. A distinction between judgments of frequency and of desirability as determinants of response. Educational and Psychological Measurement, 1969, 29, 273-293.

* Jackson, D. N., & Minton, H. L. A forced-choice adjective preference scale for personality assessment. Psychological Reports, 1963, 12(2), 515-520.

Forced-Choice Item, Reliability, Response Bias, Paired Comparison Items, Check-List, Personality Measures — 2, 12, 14

Psych. Abst., 38, #4303 — A-H

Jackson, D. N., & Morf, M. E. An analysis of two response styles: true responding and item endorsement. Educational and Psychological Measurement, 1972, 32(2), 329-53.

Jackson, R. M., & Rothney, J. W. A comparative study of the mailed questionnaire and the interview in follow-up studies. Personnel and Guidance Journal, 1961, 39, 569-571.

Jackson, D. N., et al. An evaluation of forced-choice and true-false formats in personality assessment. Princeton, New Jersey: Educational Testing Service, 1971. Report No. RB-71-67.

N/A	N/A
N/A	A-H

Jackson, D. N., et al. An evaluation of forced-choice and true-false item formats in personality assessment. Journal of Research in Personality, 1973, 7(1), 21-30.

Jacobs, S. S. An experimental analysis of answer-changing behavior on objective tests (ERIC Document Reproduction Service, ED 048 345). Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, 1971.

Achievement Measures	18
ERIC Document Reproduction Service, ED 048 345	A-N

Jacobs, S. S. Answer changing on objective tests: Some implications for test validity. Educational and Psychological Measurement, 1972, 32(4), 1039-1044.

Achievement Measures	18
ERIC Document Reproduction Service, EJ 069 060	A-N

Jacobs, T. O., et al. A guide for developing questionnaire items. Alexandria, Virginia: Human Resources Research Organization, 1970. Report No. HUMRRO-RBP-D4-70-1.

Textbook	17
ORA	R-H

Jacobson, E. A. A comparison of the effects of instructions and models upon interview behavior of high-dependent and low-dependent subjects. Dissertation Abstracts, 1969, 29(9-B), 3485.

* Jacoby, J., & Matell, M. Three-point Likert scales are good enough. Journal of Marketing Research, 1971, 8, 495-500.

Instrument Format, Rating Scales, Literature Review, Response Alternatives, Reliability, Validity	3a
ORA	R-H

* Jahn, J. A. Some further contributions to Guttman's theory of scale analysis. American Sociological Review, 1951, 16, 233-239.

Data Analysis, Scaling	15
Psych. Abst., 27, #1069	A-N

Jahoda, M., Deutsch, M., & Cook, S. W. Data collection, the questionnaire and interview approach. In Research methods in social relations with especial reference to predjudice, New York: Dryden Press, 1951. Pp. 151-208.

Jain, S. K. Concept of soldiers' efficiency. Psychological Researches, 1966, 1(1-2), 15-23.

Jakobovits, L. A. The affect of symbols: Towards the development of a cross-cultural graphic differential. International Journal of Symbology, 1969, 1(1), 28-52.

* Jarrett, R. F., & Sherriffs, A. C. Forced-choice versus permissive techniques in obtaining responses to attitude questionnaires. *Journal of General Psychology*, 1956, *55*, 203-206.

Forced Choice Items, Investigator Error — 7

Psych. Abst., *33*, #8127 — A-H

* Jenkins, J. G. Characteristics of the question as determinants of dependability. *Journal of Consulting Psychology*, 1941, *5*, 164-169.

Clarity, Investigator Error, Instrument Format — 3g, 13, 4

ORA — R-M

* Jenkins, J. J., Russell, W. A., & Suci, J. An atlas of semantic profiles for 360 words. *American Journal of Psychology*, 1958, *71*, 688-699.

Semantic Differential Items, Reliability — 14

Report Summary — R-M

* Jensen, J. A., & Schmitt, J. A. The influence of test title on test response. *Journal of Educational Measurement*, 1970, *7*, 241-246.

Investigator Error, Personality Measures, Respondent's Motivation — 13, 7, 11

Journal of Educational Measurement, *7*, p. 241 — R-NA

* Jensen, M. B. An evaluation of three methods of presenting true-false examinations. *School and Society*, 1930, *32*, 675-677.

Instrument Format, True-False Items, Reliability — 3g, 7

ORA — R-H

Johnson, D. M. Confidence and speed in the two-category judgment. *Archives of Psychology*, New York, 1939, No. 241.

Psychophysical Measures, Instrument Length — 5

ORA — R-N

* Johnson, D. M. Reanalysis of experimental halo effects. *Journal of Applied Psychology*, 1963, *47*, 46-47.

Data Analysis, Investigator Error, Rating Scales 7, 12

Journal of Applied Psychology, *47*, 46 (Rev.) R-H

* Johnson, D. M., & Vidulich, R. N. Experimental manipulation of the halo effect. *Journal of Applied Psychology*, 1956, *40*, 130-134.

Investigator Error 12

Psych. Abst., *31*, #5792 A-M

* Johnson, K. K. Company interviewers rate job applicants. *Personnel and Guidance Journal*, 1958, *36*, 422-424.

Interviews, Investigator Error 9, 13

Psych. Abst., *33*, #7000 A-N

Johnson, R. H., & Bond, G. L. Reading ease of commonly used tests. *Journal of Applied Psychology*, 1950, *34*, 319-324.

Clarity 4,9

Psych. Abst., *26*, #299 (Rev.) A-M

Jones, E. E., & Sigall, H. The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 1971, *76*, 349-364.

Jones, L. V., & Bock, R. D. *The Measurement and Prediction of Judgment and Choice*. San Francisco: Holden-Day, 1968.

Jones, L. V., & Jeffrey, T. E. *Development of suitable rating scales for measuring the subjective reactions of troops using QM items under actual field test conditions*. Chapel Hill, N. C.: University of North Carolina, Psychometric Laboratory, 1959.

* Jones, L. V., & Thurstone, L. L. The psychophysics of semantics: An experimental investigation. Journal of Applied Psychology, 1955, 39, 31-36.

Adjectives 6

ORA R-H

Jones, L. V., Thurston, L. L., & Maston, R. Measurement of food preference and prediction of consumer choice. QM Food and Container Institute, 1953. Project No. 7-84-15-007.

* Jones, M. B. The deliberate use of a set to 'fake' in personality questionnaires. Pensacola, Fla.: Naval School of Aviation Medicine, 1959. Report No. 29.

Response Bias Personality Measures 12

Jones, 1958 A-M

* Jones, R. R. Differences in response consistency and subject's preferences for three personality inventory response formats. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 3, 247-248. (Summary)

Multiple-Choice Items, Response Alternatives, Respondent's Motivation, Reliability, True-False Items 2, 3a, 11

ORA R-H

Jones, S. Process testing - an attempt to analyze reasons for students' responses to test questions. Journal of Educational Research, 1953, 46, 525-534.

Achievement Measures 18

Psych. Abst., 28, #3292 A-N

* Jury, P. A. The relation of sample demographic characteristics to job satisfaction. University of Minnesota, Technical Report No. 9001. 1971.

Investigator Error 9

DDC, # AD 724 809 A-M

Kahn, R. L., & Cannell, C. F. *The dynamics of interviewing*. New York: John Wiley, 1957.

Interviews	13
Bauman, Rogers, and Weiss, 1971	A-M

Kahn, R. L., & Cannell, C. F. Interviewing: I. Social research. in *International Encyclopedia of the Social Sciences*, D. L. Sills (Ed.), New York: Macmillan and Free Press, 1968, *8*, 149-161.

Kamenetzky, J., Burgess, G. G., & Rowan, T. The relative effectiveness of four attitude assessment techniques in predicting a criterion. *Educational and Psychological Measurement*, 1956, *16*, 187-197.

Rating Scales, Projective Items	2
Psych. Abst., *31*, #5101	A-M

Kamfer, L. The utility of a buddy rating procedure as opposed to a sociometric test for the identification of military leaders. *Psychol. Afr*., 1962, *9*, 37-43.

Rating Scales	1
Psych. Abst., *37*, #5718	A-M

* Kane, R. B. *Reducing proximity error in administering the semantic differential* (ERIC Document Reproduction Service, ED 032 761). Lafayette, Indiana: Purdue University, 1968. Contract OEC-0-070189-2508.

Instrument Format, Semantic Differential Items	3c, 2
ERIC Document Reproduction Service, ED 032 761	A-H

Kane, R. B. Computer generation of semantic differential (SD) questionnaires. *Educational and Psychological Measurement*, 1969, *29*(1), 191-192.

Kane, R. B. Minimizing order effects in the semantic differential. Educational and Psychological Measurement, 1971, 31(1), 137-144.

Semantic Differential Items, Response Alternatives, Instrument Format — 3b, 3c

ORA — R-H

Kapoor, S. D. A comparative study of the personality questionnaire items presented in the first and second person. Manas, 1963, 10, 35-44.

Question Stem, Investigator Error, Personality Measures — 3g

Psych. Abst., 38, #8496 — A-H

* Karr, C. A comparison of EPPS scores obtained from the standard forced-choice procedure and a rating-scale procedure. Dissertation Abstracts, 1959(a), 19, 3382-3383.

* Karr, C. Two methods for scoring self-rating scales to approximate forced-choice results. Psychological Reports, 1959 (b), 5, 773-779.

Scoring — 8

Psychological Reports, 5, p. 778 — R-M

Kashiwagi, S. Psychological rating of human fatigue. Ergonomics, 1971, 14(1), 17-21.

* Kassarjian, H. H., & Nakanishi, M. Study of selected opinion measurement techniques. Journal of Marketing Research, 1967, 4, 148-153.

Check List, Ranking, Paired Comparison Items, Response Alternatives, Rating Scales — 2, 3a

ORA — R-H

Katona, G., Mandell, L., & Schmiedeskamp, J. W. Questionnaire - Chapter 15. In Survey Research Center (Ed.), *1970 Survey of consumer finance*. Ann Arbor, Mich.: Michigan University, Survey Research Center, Institute for Social Research, 1971.

Katz, D. Do interviewers bias poll results? *Public Opinion Quarterly*, 1944 (a), *8*, 468-482.

Katz, D. The measurement of intensity. In H. Cantril et al., *Gauging public opinion*. Princeton, N. J.: Princeton University Press, 1944 (b). Pp. 51-65.

Katz, D. The interpretation of survey findings. *Journal of Social Issues*, 1946, *2*, 33-44.

Katz, D., Cartwright, D., Eldersveld, S. J., & Lee, A. M. (Eds.) *Public opinion and propaganda - A book of readings*. New York: Holt, Rinehart and Winston, 1954.

Katz, D. B., et al. *Guide to special issues and indexes to periodicals.* Special Libraries Association, 1962.

Kauffman, R. E. The open-ended and closed question: Some basic considerations. *New Scholar*, 1970, *2*, 101-118.

Kawash, M. B., & Aleamoni, L. M. Effect of personal signature on the initial rate of return of a mailed questionnaire. *Journal of Applied Psychology*, 1971, *55* (6), 589-592.

Kay, H. A new approach to projective testing in survey research. *Public Opinion Quarterly*, 1955, *19*(3), 267-278.

Keating, E., Paterson, D. G., & Stone, C. H. Validity of work histories obtained by interview. *Journal of Applied Psychology*, 1950, *34*, 6-11.

Keats, J. A. A method of treating individual differences in multi-dimensional scaling. British Journal of Statistical Pscyhology, 1964, 17, 37-50.

Data Analysis 15

British Journal of Statistical Psychology, 17, p. 37 R-N

Keats, J. A. Test theory. Annual Review of Psychology, 1967, 18, 217-238.

Keil, W. Investigations on successive item reversions in self-descriptive statements. Archiv fur Psychologie, 1971, 123(3-4), 285-323.

* Kelley, H. H., Hovland, C. I., Schwartz, M., & Abelson, R. P. The influence of judge's attitudes in three methods of scaling. Journal of Social Psychology, 1955, 42, 147-158.

Attitude Measures, Paired Comparison Items, Rating Scales, Scaling 10, 8, 2

ORA R-H

Kelley, R. F., & Stephenson, R. The semantic differential: An information source for designing patronage appeals. Journal of Marketing, 1967, 31, 43-47.

Semantic Differential Items 17

ORA R-N

Kelley, T. L. The scoring of alternative responses with reference to some criterion. Journal of Educational Psychology, 1934, 25, 504-510.

Kelley, T. L. Development of an activity preference test. USAF Personnel Training Research Center, 1957. Report No. 57-107.

Military Personnel, Preference Measures, Reliability 18

Psych. Abst., 33, #9312 A-N

Kellogg, W. N. The time of judgment in psychometric measures. American Journal of Psychology, 1931, 43, 65-86.

Kelly, E. L., Miles, C. C., & Terman, L. M. Ability to influence one's score on a typical pencil and paper test of personality. Character and Personality, 1936, 4, 206-215.

Kendall, L. M. An investigation of hypotheses regarding the way adjustment of time limits affects validity. Paper read at the 33rd meeting of the Eastern Psychological Association, Atlantic City, 1962.

N/A	N/A
N/A	T-H

Kendall, L. M. The effects of varying time limits on test validity. Educational and Psychological Measurement, 1964, 24, 789-800.

Multi-Choice Theory, Military Personnel, Investigator Error, Achievement Measures	7
Psych. Abst., 39, #7729	A-N

Kendall, M. G. Rank correlation methods. London: Griffin, 1948.

Kendall, M. G., & Smith, B. B. The problem of n rankings. Annals of Mathematical Statistics, 1939, 10, 275-287.

* Kendall, P. Conflict and mood; factors affecting stability of response. Glencoe, Ill.: Free Press, 1954.

Response Bias, Reliability, Investigator Error, Response Alternatives, Respondent's Motivation, Clarity	12, 3g, 11, 10, 4
Psych. Abst., 29, #2101	A-H

Kentucky University, Home Economics Education. Listing of evaluative and other types of instruments. Lexington, Kentucky: Kentucky University, Home Economics Education, 1962.

Keohane, J. Methods for surveying employee attitudes. Occupational psychology, 1971, 45, 217-231.

Kephart, W., & Bressler, M. Increasing the responses to mail questionnaires: A research study. Public Opinion Quarterly, 1958, 22(2), 123-132.

Kinard, A. M. Randomizing error in multiple-choice questions. Journal of Marketing, 1955, 19, 260-262.

Multiple Choice Items 7, 17, 3c

Bureau of Census R-N

King, F. W. Anonymous versus identifiable questionnaire in drug usage surveys. American Psychologist, 1970, 25, 982-985.

King, J. E. Multiple-item approach to merit rating. American Psychologist, 1949, 4, 278.

King, M. B., Jr. Reliability of the idea centered question in interview schedules. American Sociological Review, 1944, 9, 57-64.

Kingsbury, F. A. Analyzing ratings and training raters. Journal of Personnel Research, 1923, 1, 377-383.

Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. Sexual behavior in the human male. Philadelphia and London: W. B. Saunders, 1948.

Kirby, D., & Gardner, R. Norms on 208 words typically used in the assessment of ethnic stereotypes. *University of Western Ontario Research Bulletin 177*, 1971.

Kirkpatrick, C. Assumptions and methods in attitude measurement. *American Sociological Review*, 1936, *1*, 75-88.

Kish, L. *Survey sampling*. New York: John Wiley, 1965.

Klein, E. B. Stylistic components of response as related to attitude change. *Journal of Personality*, 1963, *31*, 38-51.

* Klein, S. M., Maher, J. R., & Dunnington, R. A. Differences between identified and anonymous subjects in responding to an industrial opinion survey. *Journal of Applied Psychology*, 1967, *51*, 152-160.

Response Bias, Respondent's Motivation, Anonymous Respondent, Investigator Error, Question Stem — 11, 12, 3g

Journal of Applied Psychology, *51*, p. 152 — R-H

Knoell, D. M., & DeGaugh, R. A. *A scaling technique designed to give approximations to factor scales*. USAF Personnel Training Research Center, *1957*. Research Report No. 57-21.

Knower, F. H. An inventory of public opinion pollers' interviewing problems. *International Journal of Opinion and Attitude Research*, 1951, *5*, 221-228.

Interviews, Investigator Error, Validity Reliability

Psych. Abst., *26*, #5492

Knowles, J. B. Acquiescence response set and the questionnaire measurement of personality. *British Journal of Social and Clinical Psychology*, 1963, *2*, 131-137.

Response Bias, Rating Scales, True-False Items — 12, 2

Psych. Abst., *38*, #4306 — A-H

* Knudsen, D. D., Pope, H., & Irish, D. P. Response differences to questions on sexual standards: An interview - questionnaire comparison. _Public Opinion Quarterly_, 1967, _31_, 290-297.

Interviews, Anonymous Respondent — 1, 11

Potter, Sharpe, Hendee and Clark, 1972 — R-H

Koehler, R. A. A comparison of the validities of conventional choice testing and various confidence marking procedures. _Journal of Educational Measurement_, 1971, _8_(4), 297-303.

Koehler, R. A. _Coombs' type response procedures_ (ERIC Document Reproduction Service, ED 063 338). Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, 1972.

Kogan, W. S., & Fordyce, W. E. The control of social desirability: A comparison of three different Q sorts and a check list, all composed of the same items. _Journal of Consulting Psychology_, 1962, _26_(1), 26-30.

* Kohan, S., deMille, R., & Myers, J. Two comparisons of attitude measures. _Journal of Advertising Research_, 1972, _12_, 29-34.

Close-Ended Items, Open-Ended Items, Rating Scales — 1, 2

ORA — R-M

* Komorita, S. S. Attitude content, intensity, and the neutral point on a Likert scale. _Journal of Social Psychology_, 1963, _61_, 327-334.

Data Analysis, Scaling, Attitude Measures, Scoring — 2, 8

Journal of Social Psychology, _61_, pp. 327-334 — R-M
(Rev. from rept.)

* Komorita, S. S., & Graham, W. K. Number of scale points and the reliability of scales. _Educational and Psychological Measurement_, 1965, _25_(4), 987-995.

Response Alternatives, Reliability — 3a, 8

Psych. Abst., _40_, #3576 — A-H

* Kornhauser, A. The problem of bias in opinion research. *International Journal of Opinion and Attitude Research*, 1947, *1*(4), 1-16.

Investigator Error, Response Bias 12

ORA R-M

Kornhauser, A. Constructing questionnaires and interview schedules. In Jahoda, M., Deitsch, M., & Cook, S. W. (Eds.), *Research methods in social relations*. New York: Dryden Press, 1951. Pp. 423-462.

Kornhauser, A., & Sheatsley, P. B. Questionnaire construction and interview procedure. In Sellitz, C., Jahoda, M., Deutsch, M., & Cook, S. W. (Eds.), *Research methods in social relations*. New York: Holt, Rinehart, and Winston, 1959. Appendix C, pp. 546-574.

Kotula, L. J., & Haggerty, H. R. *Research on the selection of officer candidates and cadets*. U. S. Army Personnel Research Organization, 1966. Research Report No. 1146.

Krause, M. S. The validity of ratings. *Psychological Reports*, 1960, *7*, 71-79.

Krause, M. S. Role-deviant respondent sets and resulting bias, their detection and control in the survey interviews. *Journal of Social Psychology*, 1965, *67*(1), 163-183.

Krause, M. S. Ordinal scale construction for convergent validity, object discrimination, and resolving power. *Multivariate Behavioral Research*, 1966, *1*(3), 379-385.

Krech, D., and Crutchfield, S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.

Krech, D., Crutchfield, S., and Ballachey, E. Individual in society. New York: McGraw-Hill, 1962.

* Kriedt, P. H., & Clark, K. E. "Item analysis" versus "scale analysis." Journal of Applied Psychology, 1949, 33, 114-121.

Scaling, Data Analysis, Attitude Measures 8, 14

Psych. Abst., 26, #23 A-N

Krieger, M. H. A control for social desirability in a semantic differential. British Journal of Social and Clinical Psychology, 1964, 3(2), 94-103.

Response Bias, Semantic Differential Items, Scaling 12

Psych. Abst., 39, #7600 A-M

Kroeger, H. J. The usefulness of the multiple-choice question. International Journal of Opinion and Attitude Research, 1947, 1(1), 102-105.

Krueger, W. C. F. An experimental study of certain phases of a true-false test. Journal of Educational Psychology, 1932, 23, 81-91.

* Krug, R. E. A selection set preference index. Journal of Applied Psychology, 1958, 42, 168-170.

Adjectives, Response Bias 6, 12

ORA R-H

Krug, R. E., & Moyer, K. E. An analysis of the F scale: II, Relationship to standardized personality inventories. Journal of Social Psychology, 1961, 53, 293-301.

* Krug, R. E., & Northrup, D. Judgment time for forced-choice adjective pairs. Journal of Applied Psychology, 1959, 43(6), 407-410.

Forced Choice Items, Adjectives, Personality Measures 6

ORA R-H

Kuang, H. P. A critical evaluation of the relative efficiency of three techniques in item analysis. Educational and Psychological Measurement, 1952, 12, 248-266.

* Kundu, G. A new technique for the measurement of attitude. Psychological Studies, Mysore, 1960, 5(2), 106-117.

* Kundu, G. The rationale for a new technique of attitude measurement. Indian Journal of Psychology, 1962, 37(3), 147-151.

Scaling, Response Bias 12, 14

Psych. Abst., 38, #10057 A-M

Kundu, G. A comparison of the Likert and a new technique of attitude measurement. Indian Journal of Psychology, 1972 (a), 47(3), 245-258.

NA NA

NA T-H

Kundu, G. A new technique of attitude scale construction. Behaviorametric, 1972 (b), 2(1), 13-20.

Kurt, B. Combining responses on two forms of a questionnaire with options in inverse order. Public Opinion Quarterly, 1949-50, 13(4), 688-689.

Multiple Choice Items, Data Analysis, Investigator 8,2

ORA R-N

Kuusinen, J. Affective and denotative structures of personality ratings. Journal of Personality and Social Psychology, 1969, 12(3), 181-188.

La Fave, L., & Sherif, M. Reference scales and placement of items with the own categories technique. Paper presented to American Psychological Association, Annual Meetings, St. Louis, 1962. Norman, Oklahoma: Institute of Group Relations, mimeographed.

N/A	N/A
N/A	T-H

* Landon, E. L., Jr., Order bias, the ideal rating, and the semantic differential. Journal of Marketing Research, 1971, 8, 375-378.

Instrument Format, Semantic Differential Items	3c
ORA	R-H

Lanfeld, E. S., & Saunders, D. R. Some determiners of performance on an experimental sentence arrangement-test. Journal of Clinical Psychology, 1961, 17, 238-41.

Langer, P. Social desirability and acquiescence on the SORT. Psychological Reports, 1962, 11(2), 531-534.

Lanman, R. W., & Remmers, H. H. The "preference" and "discrimination" indices in forced-choice scales. Educational and Psychological Measurement, 1954, 14, 541-551.

Lansing, J. B., & Eapen, A. T. Dealing with missing information in surveys. Journal of Marketing, 1959, 23, 22-28.

Data Analysis	18
ORA	A-NA

Lansing, J. B., Ginsberg, G. P., & Braaten, K. *An investigation of response error*. Urbana, Ill.: University of Illinois Press, 1961.

Lantz, D. L. *The effect of reinforcement of statements of pupil-teacher relationships in free response situation on Minnesota teacher attitude inventory scores*. Research Bulletin 67-28. Princeton, New Jersey: Educational Testing Service, 1967.

N/A N/A

N/A T-M

LaPiere, R. T. Attitudes versus actions. *Social Forces*, 1934, *14*, 230-237.

Question Stem 3g

Potter, Sharpe, Hendee and Clark, 1972 A-N

Lapointe, R. E., & Auclair, G. A. The use of social desirability in forced-choice methodology. *American Psychologist*, 1961, *16*, 446. (Abstract)

* Laurent, A. Effects of question length on reporting behavior in the survey interview. *Journal of the American Statistical Association*, 1972, *67* (338), 298-305.

Question Stem, Interviews 3g

ORA R-H

Laurent, A. *VI. Question length and reporting behavior in the interview preliminary investigation*. Ann Arbor, Mich.: Survey Research Center, Institute for Social Research, n.d.

Question Stem, Interviews 3g, 5

ORA R-H

Lazarsfeld, P. F. The art of asking 'Why?'. Three principles underlying the formulation of questionnaires. *National Marketing Review*, 1935, *1*(1), 26-38.

Lazarsfeld, P. F. The controversy over detailed interviews: An offer for negotiation. Public Opinion Quarterly, 1944, 8, 38-60.

Lazarsfeld, P. F., & Barton, A. Some general principles of questionnaire classification. In Lazarsfeld, P. G., & Rosebart, M. (Eds.), The Language of Social Research. Glencoe, Ill.: The Free Press, 1963. Pp. 83-92.

Lazarsfeld, P. F., Fiske, M. The "panel" as a new tool for measuring opinion. Public Opinion Quarterly, 1938, 2, 596-613.

Lazarsfeld, P. F., & Henry, N. W. Latent structure analysis. Boston: Houghton Mifflin, 1968.

Lecznar, W. B. Comparison of test items across forms. USAF PRL TDR No. 64-3.

Instrument Length, Achievement Measures, Reliability	8
Psych. Abst., 39, #7774	A-N

* Lederman, E. Effect of the response format on the differential measurement of traits in the Thorndike Dimensions of Personality Inventory (ERIC Document Reproduction Service, Ed 052 249). 1971.

Forced Choice Items, Response Bias	2, 12
ERIC Reproduction Service, ED 052 249	A-M

Ledvinka, J. Race of interviewer and the language elaboration of black interviewees. Journal of Social Issues, 1971, 27(4), 185-197.

N/A	N/A
N/A	T-H

Lee, J. M., & Symonds, P. M. New type of objective tests: A summary of investigations. Journal of Educational Psychology, 1934, 25, 161-184.

* Leftwich, W. H., & Remmers, H. H. A comparison of graphic and forced-choice ratings of teaching performance at the college and university level. Studies in Higher Education, 1962, No. 92.

Response Bias, Forced-Choice Items, Rating Scales 2, 12

Psych. Abst., 40, #7071 A-H

* Lehman, R. S. Length of reference line in similarity ratings. Perceptual and Motor Skills, 1967, 24(1), 216.

Rating Scales, Instrument Format 3g

ORA R-H

Lehmann, D. R., & Hulbert, J. Are three-point scales always good enough? Journal of Marketing Research, 1972, 9, 444-446.

Instrument Format, Response Alternatives 3a

ORA R-M

Lehnene, R. G. Assessing reliability in sample surveys. Public Opinion Quarterly, 1971-1972, 35(4), 578-592.

Lemaine, J. M. Introduction to attitude scales. Bulletin de Psychologie, 1971-1972, 25(2-4), 72-101.

Lenski, G. E., & Leggett, J. C. Caste, class and deference in the research interview. American Journal of Sociology, 1960, 65(5), 463-467.

Lentz, T. F. Reliability of opinionnaire technique studied intensively by the retest method. Journal of Social Psychology, 1934, 5, 338-364.

Lentz, T. F. Acquiescence as a factor in the measurement of personality. Psychological Bulletin, 1938, 35, 659.

* Leslie, L. L. Increasing response rates to long questionnaires. Journal of Educational Research, 1970, 63(8), 347-350.

Instrument Length, Respondent's Motivation — 5, 11

ORA — R-M

Levine, D. B., & Miller, H. P. Response variation encountered with different questionnaire forms - an experimental study of selected techniques used in agriculture marketing research. Washington, D. C.: United States Government Printing Office, 1957. Marketing Research Report No. 163.

N/A — N/A

N/A — T-H

Levine, R. S. A study of a new type of answer sheet. Statistical Report 56-23. Princeton, New Jersey: Educational Testing Service, 1956.

N/A — N/A

N/A — T-H

* Levonian, G. Reliability of personality measurement by interview survey methods. Psychological Reports, 1963, 13(2), 467-474.

Personality Measures, Interviews, Reliability, Rating Scales — 2

Psych. Abst., 38, #8460 — A-M

Lévy-Leboyer, C. La signification des omissions dans quatre tests collectifs: étude expérimentale. (The significance of omissions in four group tests.) Travail Humaine, 1955 (a), 18, 315-321.

* Lévy-Leboyer, C. Les omissions dans les tests collectifs. (Omissions in group tests.) *Travail Humaine*, 1955 (b), *18*, 96-108.

Response Bias, Respondent's Motivation 11, 12

Psych. Abst., *30*, #91 A-M

Lewin, K. *Principles of topological psychology*. New York: McGraw-Hill, 1936.

Lewis, H. B. Studies in the principles of judgments and attitudes: IV. The operation of "prestige suggestion." *Journal of Social Psychology*, 1941, *14*, 229-256.

Lewis, L. H. Acquiescence response set: Construct or artifact. *Journal of Projective Techniques*, 1968, *32*, 578-584.

* Liberty, P. G., Jr. Methodological considerations in the assessment of acquiescence in the MA and SD scale. *Journal of Consulting Psychology*, 1965, *29*(1), 37-42.

Response Bias, Question Stem, Personality Measures 12, 3g

Psych. Abst., *39*, #7776 A-M

Liberty, P. G., Lunneborg, C. E., & Atkinson, G. C. Perceptual defense, dissumulation, and response styles. *Journal of Consulting Psychology*, 1964, *28*, 529-537.

Lichenstein, E. R., Quinn, P., & Hover, G. L. Dogmatism and acquiescent response set. *Journal of Abnormal and Social Psychology*, 1961, *63*, 636-638.

Lieberman, L. R., & Walters, S. M. The one-year-to-live situation: Essay vs. inventory. *Journal of Clinical Psychology*, 1972, *28*(2), 205-209.

* Likert, R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932, *22*(140), 1-55.

Scaling	17
ORA	T-H

Likert, R., Roslow, S., & Murphy, G. A simple and reliable method of scoring the Thurstone attitude scales. *Journal of Social Psychology*, 1934, *5*, 228-238.

Lilly, R. S. *A developmental-study of the semantic differential*. Research Bulletin 65-28. Princeton, New Jersey: Educational Testing Service, 1965.

N/A	N/A
N/A	T-M

Lindgren, H. C. The incomplete sentences test as a means of course evaluation. *Educational and Psychological Measurement*, 1952, *12*, 217-225.

Lindquist, E. F. Factors determining reliability of test forms. *Journal of Educational Psychology*, 1930, *21*, 512-520.

* Lindzey, G. A note on interviewer bias. *Journal of Applied Psychology*, 1951, *35*(3), 182-184.

Interviews, Investigator Error	12
Psych. Abst., *26*, #832	R-M

Lindzey, G. Experiments: Their planning and execution. In Lindzey, G. (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley Publishing Company, 1954.

Lindzey, G., & Borgotta, E. F. Sociometric measurement. In Lindzey, G. (Ed.), Handbook of social psychology, Vol. 1. Reading, Mass.: Addison-Wesley, 1954. Chapter 11.

* Lindzey, G. E., & Guest, L. To repeat- check lists can be dangerous. Public Opinion Quarterly, 1951, 15, 355-358.

Check List, Response Bias, Response Alternatives	2, 8, 13, 3a
Psych. Abst., 26, #5494	A-H

Link, H. C. An experiment in depth interviewing on the issue of internationalism vs. isolationism. Public Opinion Quarterly, 1943, 7, 267-279.

Linton, H. B. Dependence on external influence: Correlates in perception, attitudes, and judgment. Journal of Abnormal and Social Psychology, 1955, 51, 502-507.

* Litwak, E. A classification of biased questions. American Journal of Sociology, 1956, 62(2), 182-186.

Investigator Error, Clarity, Data Analysis	13, 4, 8
American Journal of Sociology, 62, p. 182	R-M

Lloyd, B. B., & Innes, J. M. Influence of past experience on meaningfulness of concepts on a semantic differential. Psychological Reports, 1969, 24(1), 269-270.

Loevinger, J. The technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. Psychological Bulletin, 1948, 45, 507-529.

Loevinger, J. A theory of test response. Proceedings of the 1958 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1959. Pp. 36-47.

Long, B. H., Henderson, E. H., & Ziller, R. C. Self-ratings on the semantic differential: Content versus response set. Child Development, 1968, 39(2), 647-656.

Long, L. A study of the effect of preceding stimuli upon the judgment of auditory intensities. Archives of Psychology, New York, 1937, No. 209.

Longstaff, H. P., & Jurgensen, C. E. Fakability of the Jurgensen classification inventory. Journal of Applied Psychology, 1953, 37, 86-89.

Longworth, D. S. Use of a mail questionnaire. American Sociological Review, 1953, 18, 310-313.

Lord, F. Testing if two measuring procedures measure the same dimension. Psychological Bulletin, 1973, 79, 71-72.

Lord, F. M. Reliability of multiple-choice tests as a function of number of choices per item. Journal of Educational Psychology, 1944, 35, 175-180.

Lord, F. M. Do tests of the same length have the same standard errors of measurement? Educational and Psychological Measurement, 1957, 17, 510-521.

Data Analysis, Scoring	8
Psych. Abst., 33, #4704	A-N

Lord, F. M. The self-scoring flexilevel test (ERIC Document Reproduction Service, ED 042 813). Princeton, New Jersey: Educational Testing Service, 1970. Report No. RB-70-43.

Achievement Measures	18
ERIC Document Reproduction Service, ED 042 813	A-M

* Lorge, I. Gen-Like: Halo or reality. *Psychological Bulletin*, 1937, *34*, 545-546.

Response Bias, Rating Scales, Validity, Investigator Error 12

Psychological Bulletin, *34*, 545-546 (Rev.) A-N

Loveless, E. J. *Artistic preferences, conceptual thinking and intellectual attitudes*. *Final report*. Notre Dame, Indiana: Notre Dame University, 1970.

Lozar, C. C. *Measuring techniques to determine architectural satisfaction of human needs* (DAOG8136). Champaign, Illinois: OCE Construction Engineering Research Laboratory, 1973.

Lu, K. H. A measure of agreement among subjective judgments. *Educational and Psychological Measurements*, 1971, *31*(1), 75-84.

Luchins, A. S., & Luchins, E. H. On conformity with true and false communications. *Journal of Social Psychology*, 1955(a), *42*, 283-303.

Investigator Error 7

Psych. Abst., *31*, #974 A-H

* Luchins, A. S., & Luchins, E. H. Previous experience with ambiguous and non-ambiguous perceptual stimuli under various social influences. *Journal of Social Psychology*, 1955 (b), *42*, 249-270.

Investigator Error, Instrument Format, Projective Items 3g, 13, 7

Psych. Abst., *31*, #975 (Rev. from rept.) R-H

* Lucky, A. W., & Grigg, A. E. Repression-sensitization as a variable in deviant responding. *Journal of Clinical Psychology*, 1964, *20*(1), 92-93.

Personality Measures, Response Bias 10, 12

Psych. Abst., *39*, #10058 A-M

Lundberg, C. C. A transactional conception of fieldwork. *Human Organization*, 1968, *27*(1), 45-49.

Lundberg, G. A. *Social research*. New York: Longmans, Green, 1929. Chapters 6 and 9.

* Lusk, E. J. A bipolar adjective screening methodology. *Journal of Marketing Research*, 1973, *9*, 202-203.

Semantic Differential Items, Response Alternatives	14
ORA	R-M

* Lyman, H. B. A comparison of the use of scrambled and blocked items in a multi-scale school attitude inventory. *Journal of Educational Research*, 1949, *43*(4), 287-292.

Attitude Measures, Instrument Format	3c, 8
ORA	R-H

Lyman, H. B. *Test scores and what they mean*. New Jersey: Prentice-Hall, 1963.

Lynch, D. O., & Smith, B. C. *To change or not to change item responses when taking tests: Empirical evidence for test takers* (ERIC Document Reproduction Service, ED 064 324). Paper presented to the annual meeting of the American Education Research Association, Chicago, 1972.

Achievement Items	18
ERIC Document Reproduction Service, ED 064 324	A-M

Macoby, E. A., & Macoby, N. *The interview: a tool of social science*. *Handbook of social psychology I: Theory and method*. Reading, Mass.: Addison-Wesley, 1954.

Madden, J. M. Familiarity effects in evaluative judgments. USAF WADD Personnel Laboratory, 1960 (a). Technical Note No. 60-261.

N/A	N/A
N/A	T-H

Madden, J. M. A review of some literature on judgment with implications for job evaluation. USAF WADD Personnel Laboratory, 1960 (b). Technical Note No. 60-212.

N/A	N/A
N/A	T-M

* Madden, J. M. A comparison of three methods of rating scale construction. Journal of Industrial Psychology, 1964, 2(2), 43-50.

Rating Scales, Response Alternatives, Military Personnel, Reliability	3f, 14
Psych. Abst., 35, #7168 (Rev.)	A-H

* Madden, J. M., & Bourdon, R. D. Effects on judgment of variations in rating scale format. USAF PRL Technical Document Report No. 63-2, 1963.

Rating Scales, Military Personnel, Instrument Format, Response Alternatives	3b, 3c
Psych. Abst., 37, #8302	R-H

Madden, J. M., & Bourdon, R. D. Effects of variations in rating scale format on judgment. Journal of Applied Psychology, 1964, 48, 147-151.

Instrument Format, Rating Scales	3a
Journal of Applied Psychology, 48, p. 151	R-NA

* Madow, W. G. On some aspects of response error measurement. Proceedings of the Social Statistics Section, American Statistical Association, 1965, 182-192.

Investigator Error, Interviews	13
Bureau of Applied Social Research	A-H

Magill, W. H. The influence of the form of item on the validity of achievement tests. Journal of Educational Psychology, 1934, 25, 21-28.

Magnusson, D., & Ekehammer, B. Subjective confidence and interjudge agreement as functions of amount of information: A study of interview data (AD 908 101). Stockholm, Sweden: Stockholm University, Psychological Lab., 1972, Report No. 366.

N/A	N/A
N/A	T-M

Magnusson, D., & Ryder, B. Subjective confidence and interobserver agreement for interview data as functions of interview length (AD-853 753). Stockholm, Sweden: Stockholm University, Psychological Laboratories, 1967.

N/A	N/A
N/A	T-M

Magrid, F. N., et al. Mail questionnaires: Adjunct to the Interview. Public Opinion Quarterly, 1962, 26, 111-114.

* Mahler, I. Yeasayers and naysayers - a validating study. Journal of Abnormal and Social Psychology, 1962, 64(4), 317-318.

Response Bias, Validity	12
ORA	R-M

Maller, J. B. The effect of signing one's name. School and Society, 1930 31, 882-884.

Maller, J. B. Character and personality tests: A descriptive bibliography including measures of attitudes, interest, adjustment, appreciation, moral knowledge, behavior and rating scales. New York: Teachers College, Columbia University, 1936.

Malmud, R. S. Controlled versus free completion. American Journal of Psychology, 1925, 36, 401-411.

Achievement Measures, Response Alternatives	2
ORA	R-M

Malry, R. C., & Roberts, M. The testing situation as a positive source of influence. Research Bulletin 67-35. Princeton, New Jersey: Educational Testing Service, 1967.

N/A	N/A
N/A	T-M

Mandler, G. Stimulus variables and subject variables: A caution. Psychological Review, 1951, 66, 145-149.

Manfield, M. N. A new dimension in consumer analysis--The Guttman scale. Paper presented before the Marketing Research Discussion Group of the American Marketing Association (New York Chapter), 1971.

Scaling, Questionnaire Theory and Development	14, 15
Abaum, et al. (Eds.) Scientific Marketing Research. New York: Scott Foresman, 1972. (Rev. from rept.)	R-N

Manis, J. G., Brawer, M. J., Hunt, C. L., & Kercher, L. C. Validating a mental health scale. American Sociological Review, 1963, 28(1), 108-116.

Manis, M. Comment on Upshaw's "own attitude as an anchor in equal-appearing intervals." Journal of Abnormal and Social Psychology, 1964, 68(6), 689-691.

Manne, H. A., & Willing, R. C. (Eds.) *Current research techniques in military personnel assignment* (ERIC Document Reproduction Service, ED 049 445). Proceedings of Annual Conference, Military Testing Association, 1970.

Questionnaire Theory and Development, Bibliography 17

ERIC Document Reproduction Service, ED 049 445 A-H

Manning, P. K. Problems in interpreting interview data. *Sociology and Social Research*, 1967, *51*, 302-316.

Marais, H. C. Evaluation of the semantic differential as instrument for measurement of attitudes. *Psychological Reports*, 1967, *21*(2), 591-592.

Semantic Differential Items 2

Psychological Reports, *21*(2), p. 591 (Rev. from rept.) R-M

Maranell, G. M. (Ed.) *Scaling: A sourcebook for behavioral scientists*. Chicago, Ill.: Aldine Publishing Co., 1974.

Scaling Bibliography 17

Aldine Publishing Co. A-H

Marcus, A. The effect of correct response location on the difficulty level of multiple-choice questions. *Journal of Applied Psychology*, 1963, *47*(1), 48-51.

Multiple Choice Items, Response Bias, Achievement Measures, Response Alternatives 12, 3e

Psych. Abst., *37*, #7978 A-M

Margolis, C., & Porter, C. R. The relationship of message unity ('pull') to the recipient's response potential. *Journal of Applied Psychology*, 1959, *43*(6), 367-371.

Markey, S. C., Winer, B. J., & Falk, G. H. Reliability of ratings for high-ranking officers. *American Psychologist*, 1949, *4*, 300.

Marks, E. S., & Mauldin, W. P. Problems of response in enumerative surveys. *American Sociological Review*, 1950, *15*, 649-657.

Response Bias, Interviews, Investigator Error 12, 13, 17

Psych. Abst., *26*, #4739 (Rev. from rept.) R-N

Marlett, S. A. A comparison of vicarious and direct reinforcement control of verbal behavior in an interview setting. *Journal of Personality and Social Psychology*, 1970, *16*(4), 695-703.

Marley, A. A. Some probabilistic models of simple choice and ranking. *Journal of Mathematical Psychology*, 1968, *5*, 311-332.

Marquis, K. H. *An experimental study of the effects of reinforcement, question length, and reinterviews on reporting selected chronic conditions in household interviews*. Ann Arbor, Mich.: Survey Research Center, University of Michigan, 1969.

Marquis, K. H. Effects of social reinforcement on health reporting in the household interviews. *Sociometry*, 1970 (a), *33*, 203-215.

Marquis, K. H. The use of verbal reinforcement in personal interviews and its relationship to data accuracy - Section Four. In Cannell, C. F., Marquis, K. H., & Laurent, A., *Studies of interviewing methodology-A summary of research conducted for the National Center for Health Statistics*. Ann Arbor, Michigan: Univ. of Michigan, Survey Research Center, 1970(b). PH-43-68-209.

N/A N/A

N/A T-H

Marquis, K. H. Effects of race, residence, and selection of respondent on the conduct of the interview. Chapter 12. In Lansing, J. B., Withy, S. B., Wolfe, A. C., et al., *Working papers on survey research in poverty areas*. Ann Arbor, Mich.: University of Michigan, Institute for Social Research, Survey Research Center, 1971.

Marquis, K. H., & Cannell, C. F. *Effect of some experimental interviewing techniques on reporting in health interview survey - a methodological study designed to test the effectiveness of certain questionnaire designs*. U. S. National Center for Health Statistics, Division of Health Interview Statistics, Public Health Service, Publication No. 1000, Series 2, No. 41.

* Marquis, K. H., Cannell, C. F., & Laurent, A. *Reporting health events in household interviews: Effects of reinforcement, question length, and reinterviews*. Ann Arbor, Mich.: Univ. of Michigan, Institute for Social Research, Survey Research Center Vital and Health Statistics, Series 2, No. 45, 1972.

Instrument Format, Interviews, Respondents' Motivation, Question Stem	3g, 11
Psych. Abst., *48*, #09206	A-H

Marquis, K. H., Marshall, J., & Oskamp, S. Testimony validity as a function of question form, atmosphere and item difficulty. *Journal of Applied Social Psychology*, 1972, *2*(2), 167-186.

N/A	N/A
N/A	T-H

* Marquis, K. H., Marshall, J., & Oskamp, S. *Accuracy and completeness of testimony as a function of kind of question, interrogation atmosphere, and item content*. Ann Arbor, Michigan: University of Michigan, Survey Research Center, n.d.

Clarity, Interviews, Respondent's Motivation, Question Stem, Investigator Error	3g, 4, 11, 7
Bureau of Applied Social Research (rev.)	A-H

Marsh, C. J. The influence of supplementary verbal directions upon results obtained with questionnaires. Journal of Social Psychology, 1945, 21(2), 275-281.

N/A N/A

N/A T-H

* Marsh, S. E., & Perrin, F. A. C. An experimental study of the rating scale technique. Journal of Abnormal and Social Psychology, 1925, 19, 383-399.

Rating Scales 2, 3f

ORA R-H

Marshall, J., Marquis, K. H., & Oskamp, S. Effects of kind of question and atmosphere of interrogation on accuracy and completeness of testimony. Harvard Law Review, 1971, 84(7), 1620-1643.

Marso, R. N. Test item arrangement, testing time, and performance. Journal of Educational Measurement, 1970, 7, 113-118.

* Mascaro, G. F. Category-width tendencies and acceptance of attitude statements. Perceptual and Motor Skills, 1968, 27(2), 410.

Response Bias 12

ORA R-M

Masling, J. The effects of warm and cold interaction on the administration and scoring of an intelligence test. Journal of Consulting Psychology, 1959, 23, 336-341.

Investigator Error 7

Psych. Abst., 34, #4395 A-M

Mason, J., & Smith, E. M. The influence of instructions on respondent error. *Journal of Marketing Research*, 1970, *7*, 254-5.

Massarik, F., Weschler, I. R., & Tannenbaum, R. Evaluating efficiency rating systems through experiment. *Personnel Administration*, 1951, *14*(1), 42-47.

Raters, Rating Scales, Investigator Error   2, 7, 11

Psych. Abst., *28*, #1610   A-N

MASSTER, C&A Branch, P&E Division, Directorate, CHARLIE (ACCB). *Questionnaire: ACCB Handbook of Design*. C&A Br., P&E Division, Directorate, CHARLIE (ACCB), Fort Hood, Texas, 1971.

Textbook   17

ORA   R-H

MASSTER, OSD-6 P&E Division, Directorate ALPHA. *Questionnaire: OSD-6 Handbook of Design*. Fort Hood, Texas: OSD-6 P&E Division, Directorate ALPHA, 1972.

Textbook   17

ORA   R-H

* Masters, J. R. Reliability as a function of the number of categories of a summated rating scale. *Dissertation Abstracts International*, 1973, *33* (8-A), 4180-4181.

Reliability, Response Alternatives, Rating Scales   3a

Dissertation Abstracts, *33*(8-A), 4180-4181 (Rev.)   R-NA

Matarazzo, J. D., & Wiens, A. N. *The interview: Research on its anatomy and structure*. Chicago: Aldine Publishing Co., 1974.

* Matarazzo, J. D., et al. Interviewer Mm-Hmm and interviewee speech durations. Psychotherapy: Theory, Research and Practice, 1964, 1(3), 109-114.

Investigator Error, Interviews, Respondent's Motivation	12, 11
Psych. Abst., 40, #1858	A-H

* Matell, M. S. The psychometric characteristics of Likert-type rating scales consisting of two-through nineteen steps. Dissertation Abstracts International, 1970, 30(9-B), 4406.

Scaling, Respondent's Motivation, Response Alternatives, Scoring	3a, 11, 8
ORA	R-NA

* Matell, M. S., & Jacoby, J. Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. Educational and Psychological Measurement, 1971, 31(3), 657-674.

Reliability, Validity, Response Alternatives, Rating Scales	3a
ORA	R-H

* Matell, M. S., Jacoby, J. Is there an optimal number of alternatives for Likert scale items? Effects of testing time and scale properties. Journal of Applied Psychology, 1972, 56(6), 506-509.

Attitude Measures, Response Alternatives, Rating Scales	3a
Journal of Applied Psychology, 56, p.506 (Rev.)	R-NA

Mathews, C. O. The effect of printed response words upon children's answers to two-response types of tests. Journal of Educational Psychology, 1927, 18, 445-457.

* Mathews, C. O. The effect of the order of printed response words on an interest questionnaire. Journal of Educational Psychology, 1929, 20, 128-134.

Attitude Measures, Instrument Format, Response Alternatives	3b
ORA	R-M

* Matthews, J. J., Wright, C. E., & Yudowitch, K. L. Analyses of the results of the administration of three sets of descriptive adjective phrases. Palo Alto: Operations Research Associates, 1975. (Prepared for the Army Research Institute for the Behavioral and Social Sciences, Fort Hood, Texas under Contract DAHC19-74-C-0032.)

Adjectives 6
ORA R-H

Mausner, B. The effect of instructed bias on judges in a Thurstone scale construction. (Doctoral dissertation, University of Pittsburgh), Pittsburgh, Penn.: Graduate School of Public Health, 1960.

Maxwell, A. E., & Pilliner, A. E. Deriving coefficients of reliability and agreement for ratings. British Journal of Mathematical and Statistical Psychology, 1968, 21(1), 105-116.

May, M. A., & Hartshorn, H. First steps toward a scale for measuring attitudes. Journal of Educational Psychology, 1926, 17, 145-162.

Mayer, C. S., & Pratt, R. W., J. A note on nonresponse in a mail survey. Public Opinion Quarterly, 1966-67, 30, 637-646.

Maynes, E. S. The anatomy of response errors-Consumer saving. Journal of Marketing Research, 1965, 2(4), 378-387.

* Mayo, G. D. Peer ratings and halo. Educational and Psychological Measurement, 1956, 16, 317-323.

Raters 12
Psych. Abst., 32, #2783 A-H

Mays, R. J. Relationships between length of acquaintance and nature of trait rated and agreement between raters. AFPTRC Res. Bull., 1954, 54-55.

McClanahan, A. H. A Monte Carlo evaluation of RAW, PRED, PROF, and COPAN: Four techniques for capturing and clustering rater strategies. *Dissertation Abstracts International*, 1973, *33*(8-B), 3987.

McClusky, H. Y. The negative suggestion effect of the false statement in the true-false test. *Journal of Experimental Education*, 1934, *2*, 267-273.

McCollough, C., & Van Atta, L. *Introduction to descriptive statistics and correlation*. New York: McGraw-Hill, 1965.

McConochie, W. A. Comparison of traditional attitude measuring techniques with a new technique. *Dissertation Abstracts International*, 1970, *31*(2-B), 901.

* McCord, H. Discovering the "confused" respondent: A possible projective method. *Public Opinion Quarterly*, 1951, *15*, 363-366.

Response Bias, Preference Measures 2, 12

Psych. Abst., *26*, #5495 A-H

* McCormick, E. J. *Effect of amount of job information required on reliability of incumbents' check-list reports*. USAF WADD Technical Note No. 60-142, 1960.

Military Personnel, Check List, Instrument Length, Reliability 5, 14

Psych. Abst., *35*, #7165 A-M

* McCormick, E. J., & Bachus, J. A. Paired comparisons ratings. I. The effect of ratings of reductions in the number of pairs. *Journal of Applied Psychology*, 1952, *36*, 123-127.

Paired Comparison Items, Instrument Length 3a, 5, 13

Psych. Abst., *27*, #691 A-H

* McCormick, E. J., & Roberts, W. K. Paired comparison ratings: 2. The reliability of ratings based on partial pairings. Journal of Applied Psychology, 1952, 36, 188-192.

Paired Comparison Items, Reliability, Instrument Length 3a, 5, 13

Psych. Abst., 27, #3062 A-H

McCormick, L. C. A rationale for scaling unordered attributes. American Journal of Sociology, 1948, 54, 31-35.

McDonagh, E. C., & Rosenblum, L. A. A comparison of mailed questionnaires and subsequent structured interviews. Public Opinion Quarterly, 1965, 29, 131-136.

McElwain, D. W. A note on item analysis. Guidance Review, 1964, 2(2), 56-65.

* McGarvey, H. R. Anchoring effects in the absolute judgment of verbal materials. Archives de Psychologie, Geneve, 1943, No. 281.

Questionnaire Theory and Development, Rating Scales 3f, 14

ORA R-N

McGee, R. K. The relationship between response style and personality variables - I. The measurement of response acquiescence. Journal of Abnormal and Social Psychology, 1962, 64(3), 229-233.

McGinnies, E. Cross-cultural investigation of some factors in persuasion and attitude change (AD 727 625). College Park, Maryland: Maryland Univ., Dept. of Psychology, 1971. Contract No. NONR-595(21).

Questionnaire Theory and Development 18

DDC, #AD 727 625 A-N

McGinnis, D. J. Making the most of formal inventories (ERIC Document Reproduction Service, ED 026 211). Paper presented at International Reading Association Conference, Boston, Mass., 1968.

N/A	18
ERIC Document Reproduction Service, ED 026 211	A-N

McGinnis, R. Scaling interview data. American Sociological Review, 1953, 18, 514-521.

McGuire, W. J. Suspiciousness of experimenter's intent as an artifact in research. In Rosenthal, R., & Rosnow, R. (Eds.), Artifact in Social Research. New York: Academic Press, 1968.

N/A	N/A
N/A	T-H

McQuire, W. J. The nature of attitudes and attitude change. In Lindzey, G., & Aronson, E. (Eds.), The handbook of social psychology. Reading, Mass.: Addison-Wesley, 1969.

McMorris, R. F., et al. Effects of violating item construction principles. Journal of Educational Measurement, 1972, 9(4), 287-294.

McNamara, W. J. Construction and scoring of the continuity or rearrangement test. Scholastic Review, 1936, 44, 50-57.

McNamara, W. J., & Weitzman, E. The effect of choice placement on the difficulty of multiple-choice test items. Educational and Psychological Measurement, 1945, 36, 103-113.

McNeil, K. A. A validity study of the semantic differential technique. College Student Survey, 1970, 4(2), 55-59.

McNemar, Q. Opinion-attitude methodology. *Psychological Bulletin*, 1946, *43*, 289-374.

Attitude Measures, Bibliography, Literature Review, Reliability, Validity — 16

ORA — R-H

McReynolds, P. (Ed.) *Advances in psychological assessment*, Vol. I. Palo Alto, Calif.: Science and Behavior Books, 1968.

Medland, F. F. *Predictor indices for officer personnel actions- officer indices* (DAOZ9849). Arlington, Va.: OCRD Behavioral and Systems Research Laboratory, 1973.

N/A — 18

DDC (Rev.) — A-NA

Medley, D. M. The influence of item modality on the dimension measured by a test. *Journal of Experimental Education*, 1956, *24*, 303-307.

Achievement Measures — 18

Psych. Abst., *31*, #6929 — A-N

Meehl, P. E., & Hathaway, S. R. The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology*, 1946, *30*, 525-564.

* Mehling, R. A simple test for measuring intensity of attitudes. *Public Opinion Quarterly*, 1959, *23*(4), 576-578.

Semantic Differential Items, Attitude Measurements, Rating Scales — 8, 2

Psych. Abst., *37*, #1070 — A-M

Meisels, M., & Ford, L. H., Jr. Social desirability response set and semantic differential evaluative judgments. *Journal of Social Psychology*, 1969, *78*(1), 45-54.

Meister, D., & Rabideau, G. *Human factors evaluation in systems development*. New York: John Wiley, 1965.

Mendelsohn, M., & Linden, J. *Development of an atypical response scale* (ERIC Document Reproduction Service, ED 059 244). Paper presented at the Annual Meeting of the Midwestern Psychological Association, Detroit, Michigan, 1971.

Questionnaire Theory and Development 18

ERIC Document Reproduction Service, ED 059 244 A-N

Menefee, S. C. The effect of stereotyped words on political judgments. *American Sociological Review*, 1936, *1*(1), 614-621.

Menny, J. W., & Tolsma, R. J. A discrimination index for items in instruments using group responses. *Journal of Educational Measurement*, 1971, *8* (1), 5-7.

* Merenda, P. F., & Clarke, W. V. Forced-choice vs. free-response in personality assessment. *Psychological Reports*, 1963, *13*(1), 159-169.

Forced-Choice Items, Personality Measures, Check-List, Respondent's Motivation 2, 11

Psych. Abst., *38*, #6089 A-H

Merrens, M. Generality and stability of extreme response styles. *Psychological Reports*, 1970, *27*, 802.

Merton, R. K. Fact and factitiousness in ethnic opinionnaires. *American Sociological Review*, 1940, *5*, 13-28.

Merton, R. K., Fiske, M., & Kendall, P. L. *The focused interview*. Glencoe, Ill.: The Free Press, 1956.

Messick, S. Separate set and content scores for personality and attitude scales. Research Bulletin 61-16. Princeton, New Jersey: Educational Testing Service, 1961.

Messick, S. Response style and content measures from personality inventories. Educational and Psychological Measurement, 1962, 22, 41-56.

Messick, S. J. Response sets. Research Memorandum 64-8. Princeton, New Jersey: Educational Testing Service, 1964.

N/A	N/A
N/A	T-M

Messick, S., & Fritzky, F. J. Dimensions of analytic attitude in cognition and personality. Journal of Personality, 1963, 31, 346-370.

Messick, S., & Kogan, N. Differentiation and compartmentalization in object-sorting measures of categorizing style. Perceptual and Motor Skills, 1963, 16, 47-51.

Messick, S., & Ross, J. (Eds.) Measurement in personality and cognition. New York: John Wiley, 1962.

Metfessel, N. S., & Sax, G. Systematic biases in the keying of correct responses on certain standardized tests. Educational and Psychological Measurement, 1958, 18, 787-790.

Investigator Error, True-False Items, Response Alternatives	13, 3c
Psych. Abst., 34, #152	A-M

Metzner, C. A. An application of scaling to questionnaire construction. Journal of the American Statistical Association, 1950, 45 (249), 112-118.

Clarity, Paired Comparison Items, Question Stem, Scaling	4, 14, 3g
Bureau of Census, #7111009601 (Rev	R-M

* Metzner, H., & Mann, F. A limited comparison of two methods of data collection: The fixed alternative questionnaire and the open-ended interview. *American Sociological Review*, 1952, *17* (4), 486-491.

Interviews, Attitude Measures, Respondent's Motivation, Anonymous Respondent	1, 9, 11
ORA	R-M

* Metzner, H., & Mann, F. Effects of grouping related questions in questionnaires. *Public Opinion Quarterly*, 1953, *17*, 136-141.

Instrument Format	3c
Psych. Abst., *28*, #2467 (Rev. from rept.)	R-H

Meyer, G. The choice of questions on essay examinations. *Journal of Educational Psychology*, 1939, *30*, 161-171.

Michael, J. J. The reliability of a multiple-choice examination under various test-taking instruction. *Journal of Educational Measurement*, 1968, *5*, 307-314.

Micklin, M., & Durbin, M. Syntactic dimensions of attitude scaling techniques: Sources of variation and bias. *Sociometry*, 1969, *32*(2), 194-206.

Micko, H. C., & Fischer, W. The metric of multidimensional psychological spaces as a function of the differential attention to subjective attributes. *Psychometrika*, 1970, *35*, 199-227.

* Miklich, D. R. Item characteristics and agreement-disagreement response set. *Dissertation Abstracts*, 1966, *26*(10), 6210.

Response Bias, Clarity, Question Stem	12, 4, 3g
Dissertation Abstracts, *26*(10), #6210	R-H

Milhalland, J. E. The scale values of statements of attitude toward the law after a 25-year interval. *Journal of Abnormal and Social Psychology*, 1958, *56*, 131-132.

Milhalland, J. E. Theory and techniques of assessment. *Annual Review of Psychology*, 1965, *15*, 311-346.

Miller, D. C. *Handbook of research design and social measurement,* (2nd ed.) New York: David McKay, 1970.

Miller, G. G. *Development of improved course evaluation methodology for technical training* (DF220180). Lowry AFB, Colorado: Air Force Human Resources Laboratories, June 1973.

Attitude Measures, Military Personnel 14, 9

DDC (Rev.) A-NA

Miller, I., & Minor, F. J. Influence of multiple-choice answer form design on answer-marking performance. *Journal of Applied Psychology*, 1963, *47*(6), 374-379.

Multiple Choice Items, Instrument Format 2, 3g, 1

Psych. Abst., *38*, #6646 A-M

Miller, N., Doob, A. N., Butler, D. C., & Marlowe, D. The tendency to agree: Situational determinants and social desirability. *Journal of Experimental Research in Personality*, 1965, *1*, 78-83.

Miller, S., et al. Use of the semantic differential in the study of motivation. *Psychological Reports*, 1971, *29*(3), 1279-1282.

Mills, J. D. Do effective research with a good questionnaire. *Industrial Marketing*, 1972, *57*, 54.

Milne, D. W. A comparison of scaling procedures and their evaluation as estimates of external criteria. Dissertation Abstracts, 1966, 26(9), 5542.

Mindak, W. A. A new technique for measuring advertising effectiveness. Journal of Marketing, 1956, 20, 367-378.

Semantic Differential Items	17
ORA	R-N

Mindak, W. A. Fitting the semantic differential to the marketing problem Journal of Marketing, 1961, 25, 28-33.

Semantic Differential Items	15
ORA	R-N

* Miron, M. S. The influence of instruction modification upon test-retest reliabilities of the semantic differential. Educational and Psychological Measurement, 1961, 21, 883-893.

Reliability, Semantic Differential Items, Investigator error	7, 11
Educational and Psychological Measurement, 21, p. 892	R-H

* Mittelsteadt, R. A. Semantic properties of selected evaluative adjectives: Other evidence. Journal of Marketing Research, 1971, 8, 236-237.

Adjectives	6
ORA	R-H

Modern Army Selected Systems Test, Evaluation, and Review (MASSTER). MASSTER Test Officer's Planning Manual. Fort Hood, Texas: Headquarters MASSTER, June, 1974.

* Mogar, R. E. Three versions of the F Scale and performance on the Semantic Differential. *Journal of Abnormal and Social Psychology*, 1960, *60*, 262-265.

Personality Measures, Response Bias, Semantic Differential Items 10

Psych. Abst., *34*, #7648 A-N

Mollenkopf, W. B. Time limits and behavior of test takers. *Educational and Psychological Measurement*, 1960, *20*, 223-230.

Instrument Length, Investigator Error, Achievement Measures 7

Psych. Abst., *35*, #7089 (Rev. from rept.) A-N

Moore, J. C., & Rubin, Z. Assessment of subjects' suspicions. *Journal of Personality and Social Psychology*, 1971, *17*(2), 163-170.

Moore, M. Moderator effects of ambivalence in attitude measurement. Paper presented to the Midwestern Psychological Association, Cleveland, 1972.

Moore, M. Ambivalence in attitude measurement. *Educational and Psychological Measurement*, 1973, *33*(2), 481-483.

Mordkoff, A. M. Functional vs nominal antonymy in semantic differential scales. *Psychological Reports*, 1965, *16*(3, Part 1), 691-692.

Morgan, R. Interviewer introspection in 'bias'. *Public Opinion Quarterly*, 1947, *11*(4), 615-616.

Morton, M. A., Hoyt, W. G., & Burke, L. K. A new type of test answer sheet. *American Psychologist*, 1955, *10*, 572.

Moscovici, S. Attitudes and opinions. *Annual Review of Psychology*, 1963, *14*, 231-260.

Mosel, J. N., & Cozan, L. W. The accuracy of application blank work histories. *Journal of Applied Psychology*, 1952, *36*, 365-369.

Mosel, J. N., & Goheen, H. W. The employment recommendation questionnaire: III. Validity of different types of references. *Personnel Psychology*, 1959 *12*, 469-477.

Moser, C. A. Interview bias. *Review of the International Statistical Institute*, 1951, *19*(1), 28-40.

Moser, C. A. *Survey methods in social investigation*. London: Heinemann, 1958.

Bibliography	16
Psych. Abst., *33*, #10.138	A-H

* Mosier, C. I. A modification of the method of successive intervals. *Psychometrika*, 1940, *5*, 101-107.

Adjectives, Scaling, Reliability	6, 14, 8, 2
ORA	R-H

* Mosier, C. I. A psychometric study of meaning. *Journal of Social Psychology*, 1941 (a), *13*, 123-140.

Adjectives	6
ORA	R-H

* Mosier, C. I. *Tables from a quantitative study of meaning*. (Privately issued.) 1941 (b).

Adjectives	6
N/A	T-H

Mosier, C. I., & McQuitty, J. V. Methods of item validation and abacs for item-test correlation and critical ratio of upper-lower difference. Psychometrika, 1940, 5, 57-65.

* Mosier, C. I., & Price, H. G. The arrangement of choices in M-C questions and a scheme for randomizing choices. Educational and Psychological Measurement, 1945, 5, 379-382.

Multiple Choice Items, Response Alternatives 3b, 12

ORA R-M

Moskowitz, H. R. Studies of validating evaluation panels: Test environments data, development of new measurement techniques for food acceptance (DAOD4177). Natick, Mass.: Army Natick Laboratories, 1973.

N/A N/A

N/A R-NA

Mosteller, F., Bush, R. R., & Green, B. F. Selected quantitative techniques and attitude measurement. Reading, Mass.: Addison-Wesley, 1970.

Mount, E. Communications barriers and the reference question-problem of inquiries not asking the reference questions they actually should pose to obtain information they need. Special Librarian, 1966, 57, 575-578.

Mudd, S. A. The treatment of handling-qualities rating data. Human Factors, 1969, 11(4), 321-330.

Mueller, D. J. A technique for the utilization of items with highly skewed response distributions in personality scaling. Journal of Experimental Education, 1972, 40 4), 62-64.

Mullins, C. J., Massey, I. H., & Riederich, L. D. Reasons for air force enlistment. Technical Report No. 68-101, USAF AFHRL, 1968.

* Murphy, E. F., Bailey, R. M., & Covell, M. R. Observations on methods to determine food palatability and comparative freezing quality of certain new strawberry varieties. Food Technology, 1954, 8, 113-116.

Ranking, Rating Scales 2

Psych. Abst., 29, #3266 (Rev.) A-H

Murphy, G., & Likert, R. Public opinion and the individual. New York: Harper, 1938.

Murphy, G., Murphy, L. B., & Newcomb, T. M. Experimental social psychology. New York: Harper, 1937.

Murphy, R. W. How and where to look it up. New York: McGraw-Hill, 1959.

* Muscio, B. The influence of the form of a question. British Journal of Psychology, 1916, 8, 351-389.

Instrument Format 3g

ORA R-H

Myers, A. E. Judgments of variability in two response modes. Research Bulletin 64-40. Princeton, New Jersey: Educational Testing Service, 1964.

N/A N/A

N/A T-M

Myers, C. T. A note on the standard length of a test. Psychometrika, 1961, 26, 443-446.

* Myers, C. T. The relationship between item difficulty and test validity and reliability. Educational and Psychological Measurement, 1962, 22 (3), 565-571.

Validity, Reliability, Question Stem 8

Psych. Abst., 37, #4936 A-N

Myers, J. H. Finding determinant attributes. Journal of Advertising Research, 10(6), 9-12.

Data Analysis	8
ORA	R-N

Myers, J. H., & Alpert, M. I. Determining Buying Attitudes: Meaning and measurement. Journal of Marketing, 1968, 32, 13-20.

Attitude Measures, Questionnaire Theory and Development	15
ORA	R-N

* Myers, J. H., & Hang, A. F. How a preliminary letter affects mail survey returns and costs. Journal of Advertising Research, 1967, 9(3), 37-40.

Respondent's Motivation	11
ORA	R-M

* Myers, J. H., & Warner, W. G. Semantic properties of selected evaluation adjectives. Journal of Marketing Research, 1968, 5, 409-412.

Adjectives	6
ORA	R-H

Naidoo, J. C. An inquiry into the structure of attitudes and behavior: A validation study (AD 647 210). Urbana, Illinois: University of Illinois, Group Effectiveness Research, 1966. Report No. TR-38.

Questionnaire Theory and Development	18
DDC, #647210	A-N

Nakamura, C. Y. Salience of norms and order of questionnaire items: Their effect on responses to the items. Journal of Abnormal and Social Psychology, 1959, 59, 139-142.

Instrument Format, Investigator Error	3c, 12
Psych. Abst., 34, #4199	A-M

Namaksy, J., Jr., Bonertz, D., Lyon, L. B., & Schuster, E. P. *Cost/effectiveness of Air Force all-volunteer force programs* (AD-920 615L). Maxwell AFB, Ala.: Air University, 1974. Report No. AU-1930-74.

N/A	N/A
N/A	T-M

Namias, J. Measuring variation in interviewer performance. *Journal of Advertising Research*, 1966, *6*(1), 8-12.

Interviews, Investigator Error	13
Bureau of Census, #7111012301	A-N

Nash, A. N. Modification of forced-choice format for use in personal selection and appraisal. *Psychological Reports*, 1971, *28*(1), 108-110.

Forced Choice Items, Response Alternatives	14, 3a
ORA	R-H

Nathan, E. D. Art of asking questions. *Personnel*, 1966 (a), *3*, 63-71.

Nathan, E. D. Asking questions that get results. *Supervisory Management*, 1966 (b), *11*, 4-8.

Attitude Measures	3g
ORA	R-N

National Education Association, Research Division. The questionnaire. Washington, D. C.: National Education Association, Research Division, *NEA Research Bulletin*, 1930, *13*(1), 1-15.

Naval School of Aviation Medicine. *The deliberate use of a set to "fake" in personality questionnaires*. Pensacola, Florida: Naval School of Aviation Medicine. Report No. 29, Project NM 14 12 11 Subtask 1.

Needham, J. G. The time-error in comparison judgments. Psychological Bulletin, 1934, 31, 229-243.

Neeley, T. E. A study of error in the interview. New York: Columbia Univ., 1937. (Unpublished doctoral dissertation), (S N29 1937).

N/A N/A

N/A T-H

* Neidt, C. O., & Merrill, W. R. Relative effectiveness of two types of response to items of a scale on attitudes toward education. Journal of Educational Psychology, 1951, 42, 432-436.

Attitude Measures, Rating Scales, Paired Comparison Items, Validity, Reliability, Response Alternatives 2, 3a

Psych. Abst., 26, #5102 A-H

Nelson, M. J., Denny, E. C., & Coladarci, A. P. Statistics for teachers. New York: The Dryden Press, 1956.

* Newhall, S. M. Comparability of the method of single stimuli and the method of paired comparisons. American Journal of Psychology, 1954, 67, 96-103.

Paired Comparison Items, Close-Ended Items 2

Psych. Abst., 29, #63 A-H

Newman, S. Differences between early and late respondents to a mailed survey. Journal of Advertising Research, 1962, 2(2), 37-39.

Newman, S. H. Quantitative analysis of verbal evaluations. Journal of Applied Psychology, 1954, 38, 293-296.

Newman, S. H., Bobbitt, J. M., & Cameron, D. C. The reliability of the interview method in an officer candidate evaluation program. *American Psychologist*, 1946, *1*(4), 103-109.

Nias, D. K., Wilson, D. G., & Woodbridge, J. M. Test-retest results on the conservatism scale completed under conditions of anonymity and identification. *British Journal of Social and Clinical Psychology*, 1971, *10*(3), 282-283.

Nicholls, J. Some effects of testing procedure on divergent thinking. *Child Development*, 1971, *42*(5), 1647-1651.

Nisselson, H., & Woolsey, T. D. Some problems of the household interview design for the national health survey. *Journal of the American Statistical Association*, 1959, *54*, 69-86.

Nixon, J. E. The mechanics of questionnaire construction. *Journal of Educational Research*, 1954, *47*(7), 481-487.

Noelle-Heumann, E. Wanted: Rules for wording structured questionnaires. *Public Opinion Quarterly*, 1970, *34*, 191-201.

Nordlie, P. G. *Further studies of attitude measurement by a word association technique* (AD 3 308). Baltimore, Maryland: University of Maryland, Dental School, 1952. Report No. TR-17.

Norman, R. P. Extreme response tendency as a function of emotional adjustment and stimulus ambiguity. *Journal of Consulting and Clinical Psychology*, 1969, *33*, 406-410.

Norman, W. T. Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology*, 1963(a), *47*, 225-241.

* Norman, W. T. Relative importance of test item content. *Journal of Consulting Psychology*, 1963 (b), *27*(2), 166-174.

Validity, Personality Measures, Question Stem 8, 4

Psych. Abst., *37*, #7980 A-M

* North, W. E., & Schmid, J. A comparison of three ways of phrasing Likert type attitude items. *Journal of Experimental Education*, 1960, *29*, 95-100.

Attitude Measures, Clarity, Investigator Error,
Military Personnel, Reliability 3g

ORA R-H

Norton, J. K. The Questionnaire. *NEA Research Bulletin 8*, No. 1, 1930.

Nosanchuk, T. A., & Marchak, M. P. Pretest sensitization and attitude change. *Public Opinion Quarterly*, 1969, *33*(1), 107-111.

Nowakowska, M. A model of questionnaire responding. *Przeglad Psychologiczny*, 1970, *19*, 95-114.

Nowakowski, M. A model for answering a questionnaire item. *Psychological Bulletin*, 1971, *2* (1), 37-45.

* Nuckols, R. C. A note on pre-testing public opinion questions. *Journal of Applied Psychology*, 1953, *37*, 119-120.

Clarity 4

Psych. Abst., *28*, #730 A-M

Nunnally, J. *Psychometric Theory*. New York: McGraw-Hill, 1967.

* Nunnally, U., & Husek, T. R. The phony language examination-an approach to the measurement of response bias. Educational and Psychological Measurement, 1958, 18(2), 275-282.

Response Bias 12, 9

Bureau of Census, #7131802001 A-H

Obe, E. O. Probabilistic, elimination, weighted-choice and conventional techniques of multiple-choice testing. (Unpublished doctoral dissertation, University of Pittsburgh), Pittsburgh, Pa.: 1971.

Oda, K. A study of the construction of rating scales: I. Bulletin of the Faculty of Education, Nagoya University, 1967, 7(42), 155.

* O'Dell, R. Personal interviews or mail panels? Journal of Marketing, 1962, 26(4), 34-39.

Interviews, Open-Ended Items, Clarity, Instrument Format, Respondent's Motivation 1, 3c, 4, 12

ORA R-H

* Odesky, S. F. Handling the neutral vote in paired comparison product testing. Journal of Marketing Research, 1967, 4, 199-201.

Data Analysis, Paired Comparison Items, Preference Measures, Response Alternatives, Scoring 8, 3a

Journal of Marketing Research, 4, p. 201 R-H

* Ognibene, P. Traits affecting questionnaire response. Journal of Advertising Research, 1973, 8(4), 29-34.

Investigator Error 9

ORA R-H

O'Gorman, J. G. *Rating in performance appraisal: A bibliography*. Australian Army Psychological Research Unit Research Report, No. 4, 1972.

N/A	N/A
N/A	T-H

Ogus, J. L. *Forms design - 1958 and 1963 results*. Washington, D. C.: Bureau of the Census, (64-111(MRD)), 1964.

* O'Leary, K. D. *The effects of observer bias in field-experimental settings. Final report*. (ERIC Document Reproduction Service, ED 078086). Stony Brook, N.Y.: State University of New York, Dept. of Psychology, 1973.

Investigator Error	13
ERIC Document Reproduction Service, ED 078086	A-H

* Olson, W. C. The waiver of signature in personal reports. *Journal of Applied Psychology*, 1936, 20 (4), 443-450.

Anonymous Respondent	11
Potter, Sharpe, Hendee, and Clark, 1972	A-M

O'Neill, J. J. *A comparative study of intelligibility values: Forms A and B*. U. S. Naval School of Aviation Project Report No. NM 001 104 500.47, 1955.

Achievement Measures	18
Psych. Abst., 30, #8252	A-N

Ong. J. *The opposite-form procedure in inventory construction and research*. NYC: Vantage Press, 1965.

Oosterhof, A. C., & Glasnapp, D. R. *Comparative reliabilities of the multiple choice and true-false formats* (ERIC Document Reproduction Service, ED 064 361). Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, 1972.

Achievement Measures 18

ERIC Document Reproduction Service, ED 064 361 A-N

Operational Test and Evaluation Command, *Operational test and evaluation methodology guide*. Draft. Fort Belvoir, Va.: 1973.

Data Analysis, Questionnaire Theory and Development 17

ORA R-M

Operation Research Associates. *Field Equipment Test Methodology Study*. *Mid-Term Contract Report. Appendix H*. *Questionnaire Construction and Development Techniques*. Contract No. DAAD-05-73-Q-0551. Palo Alto, Ca.: ORA, 1973.

Textbook 17

ORA R-H

Operations Research Associates. *Operational Test Methodology Guide*. *Volume II. Techniques and Guidelines. Annex E. Questionnaire Development and Analysis Technique*. Contract No. DAAG39-73-0107. Palo Alto, Ca.: ORA, 1974.

Textbook 17

ORA R-H

Oppenheim, A. N. *Questionnaire design and attitude measurement*. New York: Basic Books, 1966.

Textbook 17

Potter, Sharpe, Hendee and Clark, 1972 (Rev.) A-H

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 1962, 17, 776-783.

Orpen, C. The wheel and the table: The relative merits of two alternative instruments for collecting semantic-type data. British Journal of Educational Psychology, 1972, 42(1), 86-87.

Instrument Format 3e

ERIC Document Reproduction Service, EJ 054524 A-M

* Osburn, B. H., Lubin, A., Loeffler, J. C., & Tye, V. M. The relative validity of forced choice and single stimulus selfdescription items. Educational and Psychological Measurement, 1954, 14, 407-417.

Forced Choice Items, Instrument Format, Validity 2

Psych. Abst., 29, #1813 (Rev. from rept.) R-H

Osburn, H. G. The effect of item stratification on errors of measurement. Educational and Psychological Measurement, 1969, 29(2), 295-301.

* Osgood, C. E. The nature and measurement of meaning. Psychological Bulletin, 1952, 49, 197-237.

Questionnaire Theory and Development, Semantic Differential Items 2, 14

Psychological Bulletin, 49, p. 231 R-M

Osgood, C. E. Method and theory in experimental psychology. New York: Oxford University Press, 1953.

Osgood, C. E. On the whys and wherefores of E, P, and A. Journal of Personality and Social Psychology, 1969, 12(3), 194-199.

Osgood, C. E., & Suci, G. J. Factor analysis of meaning. Journal of Experimental Psychology, 1955, 50, 325-338.

Semantic Differential Items, Forced Choice Items, Questionnaire Theory and Development	14
Journal of Experimental Psychology, 50, p. 338	R-M

* Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana, Ill.: University of Illinois Press, 1957.

Semantic Differential	2
ORA	R-H

Ostrom, T. M. Item construction in attitude measurement. Public Opinion Quarterly, 1971-1972, 35(4), 593-600.

Owens, W. A. The form of items and the distribution of false positive scores on a neurotic inventory. Proceedings of the Iowa Academy of Science, 1946, 53, 285-288.

Owens, W. A. Item form and 'false-positive' responses on a neurotic inventory. Journal of Clinical Psychology, 1947, 3, 264-269.

Pace, C. R. A situations test to measure social-political-economic attitudes. Journal of Social Psychology, 1939, 10, 331-344.

Pack, E. C. The effects of testing upon attitude towards the method and content of instruction. Journal of Educational Measurement, 1972, 9(2), 141-144.

Palmer, G. L. Factors in the variability of response in enumerative studies. Journal of the American Statistical Association, 1943, 38(222), 143-270.

* Paradise, L. M., & Blankenship, A. B. Depth questioning. Journal of Marketing, 1951, 15, 274-288.

Interviews, Literature Review, Questionnaire Theory and Development	1, 17
ORA	R-M

Parducci, A. Direction of shift in the judgment of a single mind. Journal of Experimental Psychology, 1956, 51, 169-178.

Psychophysical Measures	18
Psych. Abst., 31, #2378	A-N

Parducci, A. Range-frequency compromise in judgment. Psychological Monographs, 1962, 77(2, No. 565).

Parducci, A., & Perrett, L. F. Category rating scales: effects of relative spacing and frequency of stimulus values. Journal of Experimental Psychology, 1971, 89, 427-452.

N/A	N/A
N/A	T-H

* Parker, C. A., Wright, E. W., & Clark, S. G. Questions concerning the interview as a research technique. Journal of Educational Research, 1957, 51, 215-222.

Reliability, Interviews, Close-Ended Items	1
Psych. Abst., 33, #3802	A-M

Parker, G. V., & Veldman, D. J. Item factor structure of the Adjective Check List. Educational and Psychological Measurement, 1969, 29, 605-613.

Adjectives, Check List	6
ORA	R-M

Parris, H. L. A comparative study of forced-choice and check-list ratings of Air Force R.O.T.C. instructors. (Doctoral dissertation, Ohio State University) Columbus, Ohio: Graduate Division, 1951.

N/A	N/A
N/A	T-H

Parry, H. J., & Crossley, H. M. Validity of responses to survey questions. Public Opinion Quarterly, 1950, 14, 61-80.

Parsons, H. Man-machine system experiments. Baltimore: The Johns Hopkins Press, 1972.

Parten, M. B. Surveys, polls, and samples: Practical procedures. New York: Harper, 1950.

Pasanella, A. K., et al. Bibliography of test criticism (ERIC Document Reproduction Service, ED 039 395). New York: College Entrance Examination Board, 1967.

Achievement Measures	18
ERIC Document Reproduction Service, ED 039 395	A-N

Patterson, A. C. The questionnaire as a means of educational research. III. The construction and administration of a questionnaire. Scottish Educational Journal, 1942, 25, 708.

Paul, L. E. The construction of interval scales for measuring the acceptability of clothing and equipment in field tests. Quartermaster R&E Field Evaluation Agency, 1960. Technical Report R-4, Project No. 07-98-05-001.

* Paull, D. Reliability of ordinal scales derived by ego-involved judges. Journal of Social Psychology, 1968, 76(1), 143-144.

Response Bias, Paired Comparison Items, Ranking	12, 2, 10
Psych. Abst., 44, #1758	A-H

Payne, S. L. Thoughts about meaningless questions. Public Opinion Quarterly, 1950 (b), 14, 687-696.

* Payne, S. L. Case study in question complexity. Public Opinion Quarterly, 1950 (a), 13, 653-658.

Clarity, Interviews, Multiple Choice Items, Investigator Error	3c, 4
Potter, Sharpe, Hendee, & Clark, 1972 (rev.)	R-H

* Payne, S. L. The art of asking questions. Princeton, N. J.: Princeton University Press, 1951 (Revised Ed., 1963).

Textbook	17
Potter, Sharpe, Hendee and Clark, 1972	A-H

* Payne, S. L. Combination of survey methods. Journal of Marketing Research, 1964, I, 61.

Interviews, Close-Ended Items	1, 17
ORA	R-N

* Payne, S. L. Are open-ended questions worth the effort? Journal of Marketing Research, 1965, 2, 417-418.

Close-Ended Items, Open-Ended Items, Interviews	2
ORA	R-H

* Peabody, D. Two components in bipolar scales: Direction and extremeness. Psychological Review, 1962, 69, 65-73.

Scaling, Scoring	8
Psychological Review, 69, p. 65-73.	R-M

Peabody, D. Attitude content and agreement set in scales of authoritarianism, dogmatism, anti-semitism and economic conservatism. Journal of Abnormal and Social Psychology, 1963, 66, 1-11.

Peabody, D. Models for estimating content and set components in attitude and personality scales. Educational and Psychological Measurement, 1964, 24, 255-69.

Peabody, D. Authoritarianism scales and response bias. Psychological Bulletin, 1966, 65, 11-23.

* Pearlin, L. I. The appeals of anonymity in questionnaire response. Public Opinion Quarterly, 1961, 25(4), 640-647.

Respondent's Motivation, Anonymous Respondent 11

Bureau of the Census, #7111033901 A-H

Pedersen, D. M., & Breglio, V. J. The correlation of two self-disclosure inventories with actual self-disclosure: A validity study. Journal of Psychology, 1968, 68, 291-298.

Peek, R. M. A comparison of four scaling techniques in the development of a behavior rating scale. Dissertation Abstracts, 1968, 29(3-B), 1177-1178.

Rating Scales, Reliability 2

Dissertation Abstracts, 29(3-B), pp. 1177-1178 (Rev.) A-M

Pelz, D. C. The influence of anonymity on expressed attitudes. Human Organization, 1959, 18 (2), 88-91.

* Penner, L., Homant, R., & Rokeach, M. Comparison of rank-order and paired-comparison methods for measuring value systems. Perceptual and Motor Skills, 1968, 27(2), 417-418.

Paired Comparison Items, Reliabilities, Ranking 2

ORA R-H

Peouts, J. H., & Rader, G. E. The influence of interviewer characteristics on the initial interview. Social Casework, 1962, 43(10), 548-552.

378

Perloff, R. Consumer analysis. Annual Review of Psychology, 1968, 19, 437-466.

Perrine, M. W. Some influences of verbal reinforcement upon reference scale formation and discrimination. Dissertation Abstracts, 1959, 19, 2172-2173.

Psychophysical Measures, Investigator Error, Respondent's Motivation 11

Dissertation Abstracts, 19, pp. 2172-2173 A-M

* Perry, D. K. Forced-choice vs. L-I-D response items in vocational interest measurement. Journal of Applied Psychology, 1955, 39, 256-262.

Forced Choice Items, Multiple Choice Items 2

Psych. Abst., 30, #4734 A-H

* Personnel Research Section, PRPB, Adjutant-General's Office. The forced-choice technique and rating scales. American Psychologist, 1946, 1, 267.

Rating Scales, Forced Choice Items 2

American Psychologist, 1, p. 267 A-M

Peryam, D. R., Polemis, B. W., Kamen, J. M., Eindhoven, J., & Pilgrim, F. J., Food Preferences of Men in the U. S. Armed Forces. Washington, D.C.: Department of the Army, Quartermaster Food and Container Institute, 1960.

N/A N/A

N/A T-M

* Peters, D. L., & McCormick, E. J. Comparative reliability of numerically anchored versus job-task anchored rating scales. Journal of Applied Psychology, 1966, 50, 92-96.

Reliability, Response Alternatives, Rating Scales 3f

Psych. Abst., 40, #4642 A-H

Peters, D. L., & Messier, V. The effects of question sequence upon objective test performance. The Alberta Journal of Educational Research, 1970, 16, 253-265.

Peters, H. N. A multiple choice supraordinality test. Journal of Clinical Psychology, 1958, 14, 416-418.

Peterson, F. A. A technique for the detection of blind checking in questionnaire research. Educational and Psychological Measurement, 1961, 21, 361-362.

Response Bias	12
ORA	R-H

Pettigrew, T. F. The measurement and correlates of category width as a cognitive variable. Journal of Personality, 1958, 26, 532-544.

Phifer, M. K. Influence of the process of discrimination on the selection of statements for an attitude scale. Public Opinion Quarterly, 1971-1972, 35(4), 601-605.

Attitude Measures, Card Sorts, Investigator Error, Scaling	3f, 10, 13
ORA	R-H

Phillips, M. Problems of questionnaire investigation. Research Quarterly, 1941, 12, 528-537.

Pickett, G. D. A comparison of translation and blank-filling as testing techniques. Available from-Subscription Department, Oxford University Press, 1968.

Pierre, K. J. Social desirability response set in clinical interviewing. Dissertation Abstracts International, 1971, 32, 1856-1857.

* Pilgrim, F. J., & Wood, K. R. Comparative sensitivity of rating scale and paired comparison methods for measuring consumer preference. _Food Technology_, 1955, _9_, 385-387.

Rating Scales, Paired Comparison Items 2

Psych. Abst., _30_, #5408 A-H

Pintner, R., & Forlano, G. The influence of attitude upon scaling of attitude items. _Journal of Social Psychology_, 1937, _8_, 39-45.

Pippette, G. L. An experiment with college questionnaires. _Journal of Marketing_, 1940, _5_, 122-124.

Check List, True-False Items, Open-Ended Items 2

ORA R-N

Politz, A. Questionnaire validity through the opinion-forming question. _The Journal of Psychology_, 1953, _36_, 11-15.

Pollack, I. Iterative techniques for unbiased rating scales. _Quarterly Journal of Experimental Psychology_, 1965, _17_(2), 139-148.

Pollaczek, P. P. A study of malingering on the CVS abbreviated individual intelligence scale. _Journal of Clinical Psychology_, 1952, _8_, 75-81.

Postman, L., & Crutchfield, R. S. The interaction of need, set, and stimulus-structure in a cognitive task. _American Journal of Psychology_, 1952, _55_, 196-217.

Postman, L., & Zimmerman, C. Intensity of attitude as a determinant of decision-time. _American Journal of Psychology_, 1945, _58_, 510-518.

Potter, D. R., Sharpe, K. M., Hendee, J. C., & Clark, R. N. Questionnaires for research: An annotated bibliography on design, construction and use. Portland, Oregon: Department of Agriculture, Pacific Northwest Forest and Range Experiment Station, 1972. Forest Service Paper PNW-140.

Bibliography 16

Potter, Sharpe, Hendee and Clark, 1972 (Rev.) A-H

* Potter, G. S., & Tinkleman, V. Anchor effects in the development of behavior rating scales. Educational and Psychological Measurement, 1970, 30 (2), 311-318.

Response Bias, Response Alternatives 3f, 12

ORA R-H

Powell, W. R. Reappraising the criteria for interpreting informal inventories. International Reading Association Conference Proceedings, 1968, 13(14), 100-109.

Prien, E. P., Barrett, G., & Svetlik, B. Uses of questionnaires in job evaluation. Journal of Industrial Psychology, 1965, 3(4), 91-94.

* Prien, E. P., Otis, J. L., Campbell, J. R., & Saleh, S. Comparison of methods of measurement of job attitudes. Journal of Industrial Psychology, 1964, 2(4), 87-97.

Close-Ended Items, Open-Ended Items 2

Psych. Abst., 40, #11553 A-M

Proctor, C. H. Variations in response errors induced by changing instructions to enumerators. Proceedings of the American Statistical Association, Social Statistics Section, 1965, 8, 51-55.

Proctor, C. H. Reliability of a Cuttman scale score. Proceedings of the American Statistical Association, Social Statistics Section, 1971, 14, 348-390.

Profession Research Associates. *A comparative study of mail questionnaire techniques*. Chicago: Putnam Pub. Co., 1959.

Proshansky, H. M. A projective method for the study of attitudes. *Journal of Abnormal and Social Psychology*, 1943, *38*, 393-395.

* Prothro, E. T. The effect of strong negative attitudes on the placement of items in a Thurstone scale. *Journal of Social Psychology*, 1955, *41*, 11-18.

Attitude Measures, Investigator Error, Scaling, Response Bias 10, 12

Journal of Social Psychology, *41*, (Rev.) R-M

* Prothro, E. T. Personal involvement and item displacement on Thurstone scales. *Journal of Social Psychology*, 1957, *45*, 191-196.

Response Bias 12

Psych. Abst., *33*, #10.179 A-H

Pugh, R. C. Empirical evidence on the application of Lord's sampling technique to Likert items. *Journal of Experimental Education*, 1971, *39*(3), 54-56.

Pusar, A. *An indirect approach to the measurement of vocational interests*. (Master's Thesis, Syracuse University), Syracuse, New York: 1972.

Pyrczak, F., Jr. *Objective evaluation of the quality of multiple-choice test items*. (Doctoral dissertation, University of Philadelphia) Philadelphia, Pa.: 1972.

Pyrczak, F. *Use of similarities between stems and keyed choices in multiple-choice items*. Paper presented at Annual Meeting of National Council for Measurement in Education, 1973.

Quereshi, M. Y. The development of the Michill Adjective Rating Scale (MARS). Journal of Clinical Psychology, 1970, 26(2), 192-196.

* Quinn, J. L. Performance appraisals: The relation between ratings and selected characteristics of the rating dyads. Dissertation Abstracts, 1967, 28B, 2172.

Response Bias, Raters 9, 12

ORA A-M

* Quinn, R. P. Conformity, personality, and the extraneous third variable-acquiescence response set. Dissertation Abstracts, 1963, 24(6), 2606.

Response Bias, Personality Measures 12

Dissertation Abstracts, 24, p. 2606 (Rev.) A-H

Quinn, S. B., & Belson, W. A. The effects of reversing the order of presentation of verbal rating scales in survey interviews. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science.

Instrument Format, Rating Scales, Response Alternatives 3b

ORA R-H

* Rainio, K. The effect of the selection situation on responses to questionnaires. Acta Psychologica, 1956, 12, 244-246.

Investigator Error 7

Psych. Abst., 32, #999 A-M

* Rambo, W. W. Equal-appearing interval scales, own-attitude, and experimental instructions: A note. Perceptual and Motor Skills, 1968, 26(3, Part I), 839-842.

Attitude Measures, Scaling, Investigator Error 7, 10

ORA R-H

Ramsey, J. O. The effect of number of categories in rating scales on precision of estimation of scale values. Psychometrika, 1973, 38(4), 513-532.

Response Alternatives	3a
Psychometrika, 38, p. 513	R-NA

Ranta, T. J. Social attitudes and response style. Educational and Psychological Measurement, 1961, 21(3), 543-557.

Rapaport, G. M., & Berg, I. A. Response sets in a multiple-choice test. Educational and Psychological Measurement, 1955, 15, 58-62.

Response Bias	12
Psych. Abst., 30, #1630	A-H

Rappeport, M. A. Comments on 'an experimental study of payments to respondents.' Public Opinion Quarterly, 1971, 35(3), 335.

* Rappard, C. A. Enkele aspecten van de juistheid der personeelsbeoordelingen. (Some aspects of the correctness of personnel ratings). Psycholgische Achergronden, 1950, 5(11/12), 98-107.

Raters	12
Psych. Abst., 26, #3635	A-H

Rau, L. Variability in response to words: An investigation of stimulus-ambiguity. American Journal of Psychology, 1958, 71, 338-349.

Ray, W. S., Hundleby, J. D., & Goldstein, D. A. Test skewness and kurtosis as functions of item parameters. Psychometrika, 1962, 27, 39-47.

Razran, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). Science, 1939, 90, 89-90.

Redl, F. On examination fear. New Era, 1936, 17, 73-75.

Reel, W. D. A study of the relationship between personal information and social skills rating. Dissertation Abstracts, 1961, 22(4), 1286.

Rees, D. W., & Copeland, N. K. The effects of serial position in checklist design. USAF WADC Technical Note 59-552, 1959.

Check List, Instrument Length, Investigator Error 3g, 13, 7

Psych. Abst., 35, #1369 A-H

Reeves, J. W. What is occupational success? Postscript 1970. Occupational Psychology, 1970, 44, 219-220.

Reich, J., & Sherif, M. Ego-involvement as a factor in attitude assessment by the own categories technique. Norman, Oklahoma: University of Oklahoma, Institute of Group Relations, 1963. (Mimeographed)

Reiling, E., & Taylor, R. A new approach to the problem of changing initial responses to multiple choice questions. Journal of Educational Measurement, 1972, 9(1), 67-70.

Remmers, H. H. The validity, reliability and halo effect of human judgments in defined situations. Psychological Bulletin, 1933, 30, 577.

Investigator Error, Validity, Reliability 12, 15

Psychological Bulletin, 30, p. 577 (Rev. from rept.) A-N

Remmers, H. H. Introduction to opinion and attitude measurement. New York: Harper, 1955.

Remmers, H. H., Gage, N. L., & Rummel, J. F. A practical introduction to measurement and evaluation. New York: Harper and Row, 1965.

Textbook	16
ORA	T-H

Remmers, H. H., & Remmers, E. M. The negative suggestion effect of true-false examination questions. Journal of Educational Psychology, 1926, 17, 52-56.

Remmers, H. H., & Silance, E. B. Generalized attitude scales. Journal of Social Psychology, 1934, 5, 298-312.

Remmers, H. H., et al. An experimental study of the relative difficulty of true-false, multiple choice, and incomplete sentence types of examination questions. Journal of Educational Research, 1923, 14, 366-371.

Remmers, H. H., et al. Studies in attitudes. Purdue University Studies in Higher Education, 1934, 25, 1-112.

N/A	N/A
N/A	T-M

Remmers, H. H., et al. Further studies in attitudes. Series II. Purdue University Studies in Higher Education, 1936, 31, 1-298.

N/A	N/A
N/A	T-M

Remmers, H. H., et al. Further studies in attitudes. Series III. *Purdue University Studies in Higher Education*, 1938, *34*, 1-151.

N/A N/A

N/A T-M

* Rennick, V. G., Grupe, J. E., Reich, E. L., & Sewell, M. R. Exploratory study of rating procedures used to analyze material received on parents' reports. *Union College Stud. Character Res*., 1954, 101-124.

Ranking, Rating Scales 2

Psych. Abst., *29*, #3751 A-M

Reuschling, T. L., & Etzel, M. J. Disappearing data source. *Business Horizons*, 1973, *16*, 17-22.

Reuss, C. F. Differences between persons responding and not responding to mail questionnaires. *American Sociological Review*, 1943, *8*, 433-438.

* Reynolds, W. H. Some empirical observations on a ten-point, poor-to-excellent scale. *Journal of Marketing Research*, 1966, *3*, 388.

Scaling, Data Analysis 2, 8

Journal of Marketing Research, 1966, *3*, p. 388 R-M

Rice, S. A. Contagious bias in the interview. *American Journal of Sociology*, 1929, *35*(3), 420-423.

Richards, E. A. A commercial application of Guttman attitude scaling techniques. *Journal of Marketing*, 1951, *21*, 166-173.

Rating Scales 17, 14

ORA R-N

Richards, P. H. Analysis of the Delphi Survey 1972-1973. AOTE National Invitational Conference Redesigning Teacher Education Pre-Conference Input, 1973.

Richardson, M. W. An empirical study of the forced-choice performance report. American Psychologist, 1949, 4, 278-279.

Richardson, M. W., & Kuder, G. F. Making a rating scale that measures. Personnel Journal, 1933, 12, 36-40.

* Richardson, S A. The use of leading questions in non-schedule interviews. Human Organization, 1960, 19(2), 86-89.

Clarity, Interviews, Investigator Error 3g, 13

Bureau of Census, #7111006701 A-H

Richardson, Dohrenwend, & Klein. Interviewing, its forms and functions. New York: Basic Books, 1965.

Riker, B. L. A comparison of methods used in attitude research. Journal of Abnormal and Social Psychology, 1944, 39, 24-42.

Riker, B. L. Comparison of attitude scales- a correction. Journal of Abnormal and Social Psychology, 1945, 40, 102-103.

Riley, J. W., Jr., & Toby, J. Sociological studies in scale analysis. New Brunswick, N. J.: Rutgers University Press, 1954.

Riley, M. W. Sociological research. New York: Harcourt, Brace and World, 1963.

Riley, M. W., & Toby J. Subject and object scales: a sociological application. American Sociological Review, 1952, 17, 287-296.

Riley, M. W., et al. Sociological studies of scale analysis. New Brunswick: Rutgers University Press, 1954.

Rimoldi, H. J. A., & Devans, J. R. Some considerations on scaling procedures. Perceptual and Motor Skills, 1960, 11, 207-213.

Rimoldi, H. J. A., & Hormaeche, M. The law of comparative judgment in successive intervals and graphic rating scale methods. Psychometrika, 1955, 20, 307-318.

Rippey, R. M. Scoring and analyzing confidence tests. (ED 060 070, MF and HC available from EDRS.)

Robertson, D. W. Source documents for the automated enlisted performance evaluation system. USN Personnel and Training, 1972. Research Memorandum SRM-72-10.

Robertson, R. J., & Malchick, D. L. The reliability of global ratings versus specific ratings. Journal of Clinical Psychology, 1968, 24(2), 256-258.

Robinson, J., & Shaver, P. Measures of social psychological attitudes. (Appendix B to measures of political attitudes). Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan, 1969.

Robinson, J. P., Rusk, J. G., & Head, K. B. Measures of political attitudes. Ann Arbor, Michigan: University of Michigan, Institute for Social Research, 1968.

Robinson, R. A. How to boost returns from mail surveys. *Printers' Ink*, 1952, *237*(6), 36.

Rodger, A. The worthwhileness of the interview. *Occupational Psychology*, London, 1952, *26*, 101-106.

Interviews	1
Psych. Abst., *27*, #701	A-N

Rodgers, F. A. Test-item format preferences of elementary-school pupils. *Elementary School Journal*, 1966, *67*(1), 45-49.

* Roeber, E. C. Comparison of seven interest inventories with respect to word usage. *Journal of Educational Research*, 1948, *42*, 8-17.

Clarity, Interest Measures	4
ORA	R-N

Roehr, G. A. Effective techniques in increasing response to mailed questionnaires. *Public Opinion Quarterly*, 1963, *27*, 299-302.

Rogers, S. The anchoring of absolute judgments. *Archives of Psychology*, New York, 1941, No. 261.

Rating Scales, Psychophysical Measures	3f
ORA	R-H

* Rohila, P., Shanhdhar, S. C., & Sharma, V. Comparison of a non-verbal interest inventory with its verbal equivalent. *Journal of Psychological Researches*, 1966, *10*(1), 32-36.

Instrument Format, Clarity	3g, 4
Psych. Abst., *40*, #7719	A-M

Rohrer, J. H. A crossvalidation of the proverbial attitudes test and the psychodynamics underlying the proverbs identifying the character and personality disorders. Washington, D. C.: Georgetown University, 1961. Contract NONR 153007.

Personality Measures 18

DDC, #262 911 A-NA

Rokeach, M. Attitude as a determinant of distortions in recall. Journal of Abnormal and Social Psychology, 1952, 47, 482-488.

Personality Measures 18

Psych. Abst., 27, #2533 A-N

Rokeach, M. The double agreement phenomenon: Three hypotheses. Psychological Review, 1963, 70, 304-309.

Rokeach, M. The role of values in public opinion research. Public Opinion Quarterly, 1968-69, 32(4), 547-559.

Rokeach, M. The nature of attitudes. International Encyclopedia of the Social Sciences, 1972.

Roman, H. S. Semantic generalization in formulation of consumer attitudes. Journal of Marketing Research, 1969, 6, 369-73.

Rommetveit, R., & Svalheim, R. Some halo effects in perception of geometrical patterns. Nordisk Psykologi, 1959, 11, 11-24.

Response Bias 12

Psych. Abst., 34, #535 A-N

Roper, E. Wording of questions for the polls. Public Opinion Quarterly, 1940, 4, 129-130.

Question Stem 3g

ORA R-N

Roper, E. Checks to increase polling accuracy. Public Opinion Quarterly, 1941, 5, 87-90.

* Rorer, L. C. The great response-style myth. Psychological Bulletin, 1965, 63(3), 129-156.

Response Bias, Question Stem 12, 3g

Bureau of Census A-M

Rosander, A. C. The Spearman-Brown formula in attitude scale construction. Journal of Experimental Psychology, 1936, 19, 486-495.

Rosander, A. C. An attitude scale based upon behavior situations. Journal of Social Psychology, 1937, 8, 3-15.

Rose, A. M. A research note on experimentation in interviewing. American Journal of Sociology, 1945, 51(6), 509-586.

Rosen, H., & Rosen, R. A. The validity of "undecided" answers in questionnaire responses. Journal of Applied Psychology, 1955, 39, 178-181.

Investigator Error 12

Potter, Sharpe, Hendee, and Clark, 1972 A-M

* Rosen, N. A. Anonymity and attitude measurement. Public Opinion Quarterly, 1960, 24, 675-79.

Anonymous Respondent 11

Potter, Sharpe, Hendee, and Clark, 1972 A-M

Rosenberg, M. J., Hovland, C. I., McGuide, W. J., Abelson, R. P., & Brehm, J. W. Attitude organization and change. New Haven: Yale University Press, 1960.

Rosenthal, I., & Ferguson, T. S. An asymptotically distribution-free multiple comparison method with application to the problem of n rankings of m objects. British Journal of Mathematical and Statistical Psychology, 1965, 18(2), 243-254.

Ranking, Data Analysis	8
Psych. Abst., 40, #6090	A-N

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

Rosenthal, R., Kermit, L. F., Freidman, C. J., & Vikan, L. L. Subjects' perception of their experimenter under conditions of experimenter bias. Perceptual and Motor Skills, 1960, 11, 325-331.

Roskam, E. E. The method of triads for nonmetric multidimensional scaling. Netherlands Journal of Psychology, 1970, 25, 404-417.

* Roslow, S., & Blankenship, A. B. Phrasing the question in consumer research. Journal of Applied Psychology, 1939, 23(5), 612-622.

Questionnaire Theory and Development, Clarity, Question Stem, Instrument Format, Investigator Error	3c, 4, 9, 12, 14, 17
Potter, Sharpe, Hendee, & Clark, 1972 (Rev.)	R-M

* Roslow, S., Wulfeck, H., & Corby, G. Consumer and opinion research: Experimental studies on the form of the question. Journal of Applied Psychology, 1940, 24, 334-346.

Clarity, Check List, Response Alternatives, Open-Ended Items	2, 3g, 4, 3a
Journal of Applied Psychology, 24, pp. 345-346.	R-M

Rosnow, R. L., et al. More on the reactive effects of pretesting in attitude research: Demand characteristics or subject commitment? Educational and Psychological Measurement, 1973, 33(1), 7-17.

Ross, I. Handling the neutral vote in product testing. Journal of Marketing Research, 1969, 6, 221-222.

* Ross, P. F. Reference groups in man-to-man job performance rating. Personnel Psychology, 1966, 19(2), 115-142.

Rating Scales, Raters 2

Psych. Abst., 40, #11555 (Rev. from rept.) R-M

Ross, R. T. Optimum orders for the presentation of pairs in the method of paired comparisons. Journal of Educational Psychology, 1934, 25, 375-382.

Paired Comparison Items, Response Alternatives 3b

ORA R-N

* Ross, R. T. A linear relationship between paired comparisons and rank order. Journal of Experimental Psychology, 1955, 50, 352-354.

Paired Comparison Items, Ranking 2

Psych. Abst., 30, #6569 A-M

Rotter, G. S. Cardinal points of agreement and disagreement. Journal of Social Psychology, 1972, 86 (2).

Royer, E. B. Some recent developments in test construction. Proceedings of the Oklahoma Academy of Science, 136(16), 107-109.

* Rozeboom, W. W., & Jones, L. V. The validity of the successive intervals method of psychometric scaling. Psychometrika, 1956, 21, 165-183.

Scaling 14

Psych. Abst., 31, #4039 A-M

Rubin, D. B. Matching to remove bias in observational studies (AD716 441) (Fiche S.R821970). Cambridge, Mass.: Harvard University, Department of Statistics, 1970.

N/A N/A

N/A T-H

Ruch, F. L. Effects of repeated interviewing on the respondent's answers. Journal of Consulting Psychology, 1941, 5(4), 179-182.

Ruch, F. L. A technique for detecting attempts to fake performance on a self-inventory type of personality test. In McNemar, W., & Merrill, M. A., Studies in Personality. New York: McGraw-Hill, 1942. Pp. 229-234.

Ruckmick, C. A. The uses and abuses of the questionnaire procedure. Journal of Applied Psychology, 1932, 16(4), 32-41.

* Rugg, D. Experiments in wording questions: II. Public Opinion Quarterly, 1941, 5, 91-92.

Investigator Error 13, 3g

ORA R-M

* Rugg, D., & Cantril, H. The wording of questions in public opinion polls. Journal of Abnormal and Social Psychology, 1942, 37, 469-495.

Multiple Choice Items, Forced Choice Items, Open-Ended Items, Instrument Format, Response Bias 2, 3a, 3g, 12

ORA R-H

* Rundquist, E. A. Form of statement in personality measurement. Journal of Educational Psychology, 1940, 31, 135-147.

Investigator Error, Personality Measures, Question Stem	3g
Journal of Educational Psychology, 31, pp. 146-147 (Rev.)	R-NA

Rundquist, E. A. The forced-choice technique and rating scales. Paper presented at the American Psychological Association, Annual Meeting, Philadelphia, Pennsylvania, 1946.

N/A	N/A
N/A	T-H

* Rundquist, E. A. Response sets-a note on consistency in taking extreme positions. Educational and Psychological Measurement, 1950, 10(1), 97-99.

Response Bias	12
ORA	R-H

Rundquist, E. A. Item and response characteristics in attitude and personality assessment: A reaction to L. G. Roper's "The great response-style myth" (AD 646 772). Psychological Bulletin, 1966, 66, 166-177.

Response Bias	12
DDC (Rev.)	A-N

Rundquist, E. A., Winer, B. J., & Falk, G. H. Follow-up validation of forced-choice items of the Army Officer Efficiency Report. American Psychologist, 1950, 5, 359.

Rupley, W. H. ERIC/RCS: The cloze procedure. Journal of Reading, 1973, 16,(6), 496-502.

Sabers, D. L., & White, G. W. The effect of differential weighting of individual item responses on predictive validity and reliability of an aptitude test. *Journal of Educational Measurement*, 1969, 6(2), 93-96.

Sadacca, R. *Dimensions of response consistency in paired comparisons*. Research Bulletin 62-14. Princeton, New Jersey: Educational Testing Service, 1962.

N/A	N/A
N/A	T-H

* Saffir, M. A. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, 2, 179-198.

Scaling, Validity, Psychophysical Measures	2,14
Psychometrika, 2, p. 179 (Rev. from rept.)	R-H

Salas, R. G. *A scale of satisfaction with Army life*. Australian Military Forces Research Report No. 8-67, III, 1967.

N/A	N/A
N/A	T-M

Saltz, E., Reece, M., & Ager, J. Studies of forced-choice methodology: Individual differences in social desirability. *Educational and Psychological Measurement*, 1962, 22(2), 365-370.

Samejima, F. A general model for free response data. *Psychometrika*, 1972, 37(1), 1-68.

Sampson, P. Using the repertory grid test. *Journal of Marketing Research*, 1972, 9, 78-81.

Samuelson, F. Agreement set and anticontent attitudes in the F scale: A reinterpretation. Journal of Abnormal and Social Psychology, 1964, 68, 338-42.

Santostefano, S. Forced-choice acts as personality measures. Dissertation Abstracts, 1958, 18, 287-288.

* Sattler, J. M. Racial experimenter effects in experimentation, testing, interviewing and psychotherapy. Psychological Bulletin, 1970, 73(2), 137-160.

Investigator Error	9
Bauman, Rogers, and Weiss, 1971	A-M

* Saunders, C., & Ward, G. A comparison of differing numbers of alternatives in two bipolar scales. Proceedings of the West Virginia Academy of Science, 1964, No. 36, 187-190.

Response Alternatives, Multiple Choice Items, Reliability	2, 3a
Psych. Abst., 40, #8271	A-H

Saupe, J. L. An empirical model for the corroboration of suspected cheating on multiple-choice tests. Educational and Psychological Measurement, 1960, 20, 475-489.

Response Bias, Investigator Error	12
Psych. Abst., 35, #3761	A-N

Sawyer, H. G. How to get at the truth with market surveys by mail. Sales Management, 1959, 82, 98.

* Sax, G., & Carr, A. An investigation of response sets on altered parallel forms. Educational and Psychological Measurement, 1962, 22(2), 371-376.

Response Bias, Instrument Format, Respondent's Motivation	12, 3c, 11
Psych. Abst., 37, #3267	A-H

* Scates, D. E., & Yoemans, A. V. Developing an objective item questionnaire to assess the market for further education among employed adults. Washington, D. C.: American Council on Education, 1950(a).

Open-ended Items	1, 5
Psych. Abst., 26, #545	A-H

* Scates, D. E., & Yoemans, A. V. The effect of questionnaire form on course requests of employed adults. Washington, D. C.: American Council on Education, Research Staff on Scientific Personnel, 1950(b).

Preference Measures, Check List, Open-Ended Items, Military Personnel, Instrument Format	1, 2, 3, 7, 14
Psych. Abst., 26, #228 (Rev. from rept.)	R-M

* Schaie, K. W. On the equivalence of questionnaire and rating data. Psychological Reports, 1962, 10(2), 521-522.

Response Bias, Data Analysis	14, 12, 9
Psych. Abst., 37, #3160	A-N

* Schaie, K. W. Scaling the scales: Use of expert judgment in improving the validity of questionnaire scales. Journal of Consulting Psychology, 1963, 27(4), 350-357.

Validity, Scoring	14
Psych. Abst., 38, #2713	A-M

Scheffé, H. An analysis of variance for paired comparisons. *Journal of the American Statistical Association*, 1952, *47*, 381-400.

Question Stem, Paired Comparison Items 3c, 2

Psych. Abst., *27*, #3148 A-N

Schendel, D. E., Wilkie, W. L., & McCann, J. M. An experimental Investigation of "Attribute Importance." In Gardner (Ed.), *Proceedings of the Second Annual Conference of the Association for Consumer Research*, College Park, Md., 1971, 243-255.

Schlinger, M. J. Cues on Q-technique. *Journal of Advertising Research*, 1969, *9*(3), 53-60.

Card Sorts, Questionnaire Theory and Development R-N

ORA 14, 17

Schlinger, M. J. Responses to advertising-varieties of liking and disliking. *Journalism Quarterly*, 1970, *47*(1), 31-40.

Schmeiser, C. B., & Whitney, D. R. *The effect of selected poor item-writing practices on test difficulty, reliability, and validity: A replication.* Paper presented at the American Educational Research Association meeting, New Orleans, Louisiana, 1973.

Schmidt, C. Semantic problems of construction and analysis of a scientific questionnaire. *Bulletin du Centre d'Etudes et Recherches Psychotechniques*, 1967, *16*(3), 265-271.

Schmidt, H. D. Uber die Zuverlässigkeit von Verhaltensbeurteilungen durch Rating Skalen. (Reliability of behavior through rating scales.) *Archiv für die gesamte Psychologie*, 1966, *118*(1-2), 47-72.

Schooler, C. A note of extreme caution on the use of Guttman scales. *American Journal of Sociology*, 1968, *74*(3), 296-301.

Schubert, S. P., & Fiske, D. W. Increase of item response consistency by prior item response. *Educational and Psychological Measurement*, 1973, *33*(1), 113-121.

* Schuessler, K. F. Item selection in scale analysis. *American Sociological Review*, 1952, *17*, 183-192.

Investigator Error, Scoring	8, 3g, 13
Psych. Abst., *27*, #4751 (Rev.)	A-H

Schulman, G. I. Asch conformity studies-conformity to experimenter and/or to the group? *Sociometry*, 1967, *30*(1), 26-40.

* Schultz, C. B. *Response set factors revealed by factor analysis of an unconfounded item pool*. Seattle, Wash.: University of Washington, 1962.

Response Bias, Personality Measures	12
ORA	R-M

Schultz, D. G. Item validity and response change under two different testing conditions. *Journal of Educational Psychology*, 1954, *45*, 36-43.

* Schuman, H. The random probe: A technique for evaluating the validity of closed questions. *American Sociological Review*, 1966, *31*(2), 218-222.

Clarity, Forced Choice Items	4
Bureau of Census, #710000801	A-H

Schuman, H. Attitudes vs. actions versus attitudes vs. attitudes. *Public Opinion Quarterly*, 1972, *36*(3), 347-354.

Schuman, H. Effects of survey question wording on survey results. Ann Arbor, Mich.: University of Michigan, School of Arts, June 1974. Contract GS-39780.

Question Stem	3g
ORA	A-H

Schuman, H., & Converse, J. M. The effects of black and white interviewers on black responses in 1968. Public Opinion Quarterly, 1971, 35(1), 44-68.

Investigator Error	9
Bauman, Rogers, and Weiss, 1971 (Rev.)	A-M

Schutz, R. E., & Foster, R. J. A factor analytic study of acquiescent and extreme response set. Educational and Psychological Measurement, 1963, 23, 435-447.

Schwartz, S. H., & Tessler, R. C. A test for a model for reducing measured attitude-behavior discrepancies. Journal of Personality and Social Psychology, 1972, 24(2), 225-236.

* Schyberger, B. W. Study of interviewer behavior. Journal of Marketing Research, 1967, 4, 32-35.

Interviews, Investigator Error	9, 12, 13
ORA	R-M

Schyberger, B. W. A case against direct questions on reading habits. Journal of Advertising Research, 6(4), 25-29.

N/A	18
ORA	A-NA

Scott, C. Research on mail surveys. Journal of the Royal Statistical Society, XXIV, Series A, General, 1961, 143-195.

* Scott, W. A. Comparative validities of forced-choice and single-stimulus tests. Psychological Bulletin, 1968, 70(4), 231-244.

Response Alternatives, Forced Choice Items 2, 3a

Psych. Abst., 44, #00117 A-H

Sears, D. O., & Abeles, R. P. Attitudes and opinions. Annual Review of Psychology, 1969, 20, 253-288.

Attitude Measures, Literature Review 16

ORA R-N

Sears, R. R. Comparison of interviews with questionnaires for measuring mothers' attitudes toward sex and aggression. Journal of Personality and Social Psychology, 1965, 2(1), 37-44.

* Seashore, R. H., & Hevner, K. A time-saving device for the construction of attitude scales. Journal of Social Psychology, 1933, 4, 366-372.

Card Sorts, Scaling, Rating Scales 2, 14

ORA R-H

Secrist, G. E. Assessment of motivational, attitudinal, and satisfaction factors related to performance in Air Force technical training. Lackland AFB, Texas: Air Force Human Resources Laboratories, June 1972.

Military Personnel A-NA

ORA 18

* Sedlacek, W. E., & Brooks, G. C. Race as an experimenter effect in racial attitude measurement (ERIC Document Reproduction Service, ED 065 525). College Park, Md.: Maryland University, Cultural Study Center, 1972.

Investigator Error 9

ERIC Document Reproduction Service, ED 065 525 A-M

Seeley, L. C., Morton, M. A., & Anderson, A. A. *Exploratory study of a Sequential Item Test*. USA PRO OCRD Technical Research Note No. 129, 1962.

Military Personnel, Instrument Format, Clarity, Instrument Length, Scoring — 14, 8, 3c, 7, 5

Psych. Abst., *38*, #945 — A-M

Seeman, W. "Subtlety" in structured tests. *Journal of Consulting Psychology*, 1952, *16*, 278-283.

Sell, T., et al. *Research methods in social relations*. New York: Holt, Rinehart, 1959.

Sellitz, C., & Cook, S. *Racial attitude as a determinant of judgments of plausibility* (AD 653 827). Boulder, Colo.: University of Colorado, Institute of Behavioral Science, 1965. Contract AF-AFOSR-436-63.

Sellitz, C., Jahoda, M., Deutsch, M., & Cook, S. W. (Eds.) *Research Methods in Social Relations*. New York: Holt, 1959.

Sessions, F. Q., et al. The development, reliability and validity of an all purpose optical scanner questionnaire form. *Public Opinion Quarterly*, 1966, *30*, 423-428.

Instrument Format — 3g

Bureau of Census, #7111034901 — A-H

* Sgan, M. L. Social reinforcement, socioeconomic status and susceptibility to experimenter influence. *Journal of Personality and Social Psychology*, 1967, *5*(2), 202-210.

Investigator Error, Respondent's Motivation — 9, 10, 11

ORA — A-H

Shaffer, L. F. Fear and courage in aerial combat. *Journal of Consulting Psychology*, 1947, 11, 137-143.

Shapiro, M. J. Discovering interviewer bias in open-ended survey responses. *Public Opinion Quarterly*, 1970, *34*(3), 412-415.

* Shapiro, S., & Eberhart, J. C. Interviewer differences in an intensive interview survey. *International Journal of Opinion and Attitude Research*, 1947, *1*(2), 1-17.

Interviews, Investigator Error 13

Bureau of the Census, #7110001801 (Rev. from rept.) R-H

Sharon, A. T. The effect of instructional conditions in producing leniency on two types of rating scales. *Dissertation Abstracts*, 1969, *29*(8-B), 3124.

Investigator Error 7, 13

ORA A-M

* Sharp, H. The mail questionnaire as a supplement to the personal interview. *American Sociological Review*, 1955, *20*(6), 718.

Interviews 1

Psych. Abst., *31*, #2753 A-N

Shaw, M. E., Worthy, M., & Blum, J. M. *Effects of number of judges upon scale values in the analysis of small group tasks*. Gainesville, Florida: Office of Naval Research, 1963. Technical Report No. 2.

N/A N/A

N/A T-M

Shaw, M. E., & Wright, J. M. *Scales for the measurement of attitudes*. New York: McGraw-Hill, 1967.

Textbook 16

ORA T-H

Sheatsley, P. B. Closed questions sometimes more valid than open. Public Opinion Quarterly, 1948, 12(1), 127.

Open-Ended Items 2, 3g

ORA R-M

Sheatsley, P. B. The influence of sub-questions on interviewers' performance. Public Opinion Quarterly, 1949, 13(2), 310-313.

Sheatsley, P. B. The harassed respondent: II. Interviewing practices. In Bogart, L. (Ed.), Current controversies in market research. Chicago, Ill.: Markham Publishing Co., 1969 (a). Pp. 39-54.

Sheatsley, P. B. Questionnaire design and question wording. Talk prepared for American Newspaper Publishers Association Research Seminar, Columbus, Ohio, July, 1969 (b).

Shellhaas, M. Motion pictures for stimulus presentation: Development and uses for opinion-attitude research interviews. Psychological Reports, 1968, 22(3, Pt. 1), 689-692.

* Shen, E. The influence of friendship upon personal ratings. Journal of Applied Psychology, 1925, 9, 66-68.

Ranking, Response Bias, Investigator Error 7, 10, 12, 13

ORA R-M

Shenck, R. L., & C. Goodman. Reactions to propaganda on both sides of a controversial issue. Public Opinion Quarterly, 1939, 3 (1), 107-112.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962 (a), 27, 125-140.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962 (b), 27, 219-246.

Sheppard, D. The adequacy of everyday quantitative expressions as measurements of qualities. *British Journal of Psychology*, 1954, *45*, 40-50.

Rating Scales, Response Alternatives 3a

Psych. Abst., *28*, #6763 (Rev. from rept.) R-M

Sherif, C. W. Established reference scales and series effects in social judgment. *Dissertation Abstracts*, *22*(6), 2083.

Response Bias, Response Alternatives 3f, 12

ORA A-H

Sherif, C. W., Sherif, M., & Nebergall, R. E. *Attitude and attitude change*. Philadelphia: W. D. Suanders, 1965.

Sherif, M. A study of some social factors in perception. *Archives of Psychology*, New York, 1935, No. 187.

* Sherif, M., & Hovland, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *Journal of Abnormal and Social Psychology*, 1953, *48*, 135-141.

Attitude Measure, Respondent's Motivation, Response Bias, Card Sorts 9, 11, 12

Psych. Abst., *28*, #732 (Rev. from rept.) R-M

Sherif, M., Taub, D., & Hovland, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, 1958, *55*, 150-155.

Investigator Error, Raters, Response Alternatives 3f

ORA R-H

Shipley, W. C., Coffin, J. I., & Hadsell, K. C. Affective distance and other factors determining reaction time in judgments of color preference. Journal of Experimental Psychology, 1945, 35(3), 206-215.

* Shipley, W. C., Norris, E. D., & Roberts, M. L. The effect of changed polarity of set on decision time of affective judgments. Journal of Experimental Psychology, 1946, 36(3), 237-243.

Paired Comparison Items, Response Bias 3g, 12

ORA R-M

Shoemaker, D. M. An application of item-examinee sampling to scaling attitudes (ERIC Documentary Reproduction Service,ED 060 026). Paper presented at the Annual Meeting of the American Research Association, Chicago, 1972.

Shor, J. Report on a verbal projective technique. Journal of Clinical Psychology, 1946, 2, 279-282.

Shukla, A. N., Sohal, T. S., & Gupta, J. P. A similitudinal study of Gausset t-test, Edwards 25-D, t HL-test, Edward and Kilpatrick's scale discrimination r-phi, bi-serial coefficient, point biserial coefficient and Guilford's phi coefficient for item analysis in the construction of summated rating scale. Indian Journal of Psychology, 1971, 46(4), 329-340.

Shulson, V., & Crawford, C. C. Experimental comparison of true-false and completion tests. Journal of Educational Psychology, 1928, 19, 580-583.

Shuttleworth, F. K. A study of questionnaire technique. Journal of Educational Psychology, 1931, 22(9), 652-658.

* Sicinski, A. Don't know answers in cross national surveys. Public Opinion Quarterly, 1970, 34, 126-129.

Response Bias 9, 12

ORA R-M

Sieber, S. D. The integration of fieldwork and survey methods. *American Journal of Sociology*, 1973, *78*, 1335-1359.

* Siegel, A. I., & Schultz, D. G. Generalized Thurstone and Guttman scales for measuring technical skills in job performance. *Journal of Applied Psychology*, 1961, *45*, 137-142.

* Siegel, A. I., & Schultz, D. G. Thurstone and Guttman scaling of job related technical skills. *Psychological Reports*, 1962, *10*, 855-861.

Scaling, Check List 12, 14

Psych. Abst., *37*, #5770 A-M

* Siegel, A. I., Schultz, D. G., & Benson, S. *Post-training performance criterion development and application: A further study into technical performance check list criteria which meet the Thurstone and Guttman scalability requirements*. Wayne, Pa.: Applied Psychological Services, 1960.

Scaling 14

Psych. Abst., *35*, #1325 A-N

Siegel, A. I., Schultz, D. G., & Fischl, M. A. Absolute scaling of job performance. *Journal of Applied Psychology*, 1968, *52*(4), 313-318.

* Siegel, L. C., & Siegel, L. Item sorts versus graphic procedure for obtaining Thurstone Scale judgments. *Journal of Applied Psychology*, 1962, *46*, 57-61.

Scaling, Card Sorts, Rating Scales 1, 2

Psych. Abst., *36*, #5 GD 57S A-H

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Siegman, A. W., & Pope, B. Ambiguity and verbal behavior in the initial interview. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, *3*, 521-522.

* Siegman, A. W., Pope, B., & Blan, T. Effects of interviewer status and direction of interviewer messages on interviewee productivity. *Proceedings of the 77th Annual Meeting, American Psychological Association*, 1969, *4* (pt.2), 541-542.

Investigator Error 13

Bauman, Rogers, and Weiss, 1971 A-M

Siegmann, P. J. A comparison of factor analysis with Guttman's scaling technique. *Dissertation Abstracts*, 1960, *20*, 3368-3369.

* Siller, J., Chipman, A. Response set paralysis: Implications for measurement and control. *Journal of Consulting Psychology*, 1963, *27*, 432-438.

Response Bias 12

Journal of Consulting Psychology, *27*, p. 432. R-H

Silverman, F. H. Intraclass correlation coefficient as an index of reliability of median scale values for sets of stimuli rated by equal-appearing intervals. *Perceptual and Motor Skills*, 1968, *26*(3, Pt. 1), 878.

Silverman, F. H. Interpretation of the correlations between sets of scale values and stimuli which have been rated for more than a single attribute. *Perceptual and Motor Skills*, 1971, *33*(2), 667-669.

* Silverman, R. E. The Edwards Personal Preference Schedule and social desirability. *Journal of Consulting Psychology*, 1957, *21*, 402-404.

Response Bias, Questionnaire Theory and Development 12, 14

Psych. Abst., *33*, #1317 A-N

Silverstein, A., & Dienstiber, R. A. Rated pleasantness and association value of 101 English nouns. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*(1), 81-86.

* Simon, R. Responses to personal and form letters in mail surveys. *Journal of Advertising Research*, 1967, *7*(1), 28-30.

Respondent's Motivation, Anonymous Respondent 11, 15

ORA R-H

* Simpson, R. H. The specific meanings of certain terms indicating different degrees of frequency. *Quarterly Journal of Speech*, 1944, *30*, 328-330.

Investigator Error, Clarity 6, 4

ORA R-H

* Sims, V. M. An evaluation of five-, ten-, and fifteen-item rearrangement tests. *Journal of Educational Psychology*, 1934, *25*, 251-257.

Achievement Measures, Instrument Format, Instrument Length, Rearrangement Items, Response Alternatives 1, 5

ORA R-N

Sims, V. M., & Knox, L. B. The reliability and validity of multiple-response tests when presented orally. *Journal of Educational Psychology*, 1932, *23*, 656-662.

Sisson, E. D. Forced choice-the new Army rating. *Personnel Psychology*, 1948, *1*, 365-381.

Sjöberg, C. A questionnaire on questionnaires. *Public Opinion Quarterly*, 1954, *18*, 423-427.

* Sjoberg, L. A study of four methods for scaling paired comparisons data. Scandinavian Journal of Psychology, 1965, 6(3), 173-185.

Scaling, Data Analysis 14

Psych. Abst., 40, #2094 A-M

Skager, R. M., Bussis, A. M., & Schultz, C. B. Comparison of information scales and like-indifferent-dislike scales as measures of interest. Psychological Reports, 1965, 16, 251-61.

Interest Measures, Achievement Measures 18

ORA A-NA

Skelley, F. R. Interviewer-appearance stereotypes as a possible source of bias. Journal of Marketing, 1954, 19(1), 74-75.

* Skindrud, K. D. An evaluation of observer bias in experimental field studies of social interaction. Final Report (ERIC Document Reproduction Service, ED 072 105). Eugene, Ore.: Oregon Research Institute, 1972.

Raters 13

ERIC Document Reproduction Service, ED 072 105 A-M

* Slater, P. The test-retest reliability of some methods of multiple comparison. British Journal of Mathematical and Statistical Psychology, 1965, 18(2), 227-242.

Ranking, Paired Comparison Items, Response Alternatives, Reliability 2, 3a

Psych. Abst., 40, #6091 A-H

Sletto, R. F. Construction of personality scales by the criterion of internal consistency. New York: Sociological Press, 1937.

* Sletto, R. F. Pretesting of questionnaires. American Sociological Review, 1940, 5, 193-200.

Attitude Measures, Instrument Format, Instrument Length, Respondent's Motivation	3c, 3g, 5, 7, 11, 14, 4
Potter, Sharpe, Hendee, & Clark, 1972 (Rev. from rept.)	R-H

* Slocum, W. L., Empey, L. T., & Swanson, H. S. Increasing response to questionnaires and structured interviews. American Sociological Review, 1956, 21(2), 221-225.

Respondent's Motivation	11
Bureau of Census, #7111003501	A-M

Slonin, M. J. Sampling. New York: Simon and Schuster, 1966.

* Small, D. O., & Campbell, D. T. The effect of acquiescence response-set upon the relationship of the F Scale and conformity. Sociometry, 1960, 23, 69-70.

Response Bias	12
Psych. Abst., 36, #2093	A-M

* Smith, D. H. Correcting for social desirability response sets in opinion-attitude survey research. Public Opinion Quarterly, 1967, 31(1), 87-94.

Data Analysis, Response Bias, Scoring	8, 12
ORA	R-H

Smith, E. M., & Mason, J. R. The influence of instructions on respondent error. Journal of Marketing Research, 1970, 7(2), 254-255.

N/A	18
ORA	R-NA

Smith, F. F. Direct validations of questionnaire data. Educational Administration and Supervision, 1935, 21, 561-575.

Smith, F. F. The relation between objectivity and validity in the arrangement of items in rank order. Journal of Applied Psychology, 1936, 20, 154-160.

Investigator Error, Preference Measures, Ranking, Raters, Reliability, Response Bias, Validity — 10, 12, 13, 14

ORA — R-H

Smith, G. H. Motivation research in advertising and marketing. New York: McGraw-Hill, 1954.

* Smith, H. L. A critique of ipsative measures with special reference to the Navy Activities Preference Blank. USN PRA Technical Bulletin No. 65-16.

Literature Review, Investigator Error, Questionnaire Theory and Development — 16, 8, 14

Psych. Abst., 39, #15270 — A-M

* Smith, H. L., & Hyman, H. The biasing effect of interviewer expectations on survey results. Public Opinion Quarterly, 1950-51, 14(3), 491-506.

Interviews, Investigator Error — 12, 13, 10, 9

Psych. Abst., 26, #2110 (Rev. from rept.) — R-H

Smith, K. An investigation of the use of "double"choice" items in testing achievement. Journal of Educational Research, 1958, 51, 387-389.

Reliability, Multiple Choice Items — 2, 3a, 10

Psych. Abst., 33, #6925 — A-N

* Smith, P. C., Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47(2), 149-155.

Rating Scales, Reliability	3f, 8
Psych. Abst., 37, #7981	A-M

Smith, P. C., Kendall, L. M., & Hulin, C. L. The measurement of satisfaction in work and retirement: A strategy for the study of attitudes. Chicago: Rand McNally, 1969.

Smith, R. G., & Nichols, H. J. Semantic differential stability as a function of meaning domain. Journal of Communication, 1973, 23(1), 64-73.

* Solomon, A. The effect of answer sheet format on test performance by culturally disadvantaged fourth grade elementary school pupils. Journal of Educational Measurement, 1971, 8, 289-290.

Instrument Format	3g
ORA	R-M

* Soueif, M. I. Extreme response sets as a measure of intolerance of ambiguity. British Journal of Psychology, 1958, 49, 329-334.

Rating Scales, Response Bias	10, 12
Psych. Abst., 33, #9993	A-N

Souren, G. Het Rosenthel effect (the Rosenthal Effect) (AD 870 528). Netherlands: Leiden Rijksuniversiteit, Psychologisch Instituut, 1969. Report No. SP-002-69.

Spaltro, E. The semantic differential as a method to measure attitudes. Contributi-Dell'istituto di Psicologia, 1967, 29, 166-235.

Speak, M. Some characteristics of respondents, partial-respondents, and non-respondents to questionnaires on job satisfaction. Occupational Psychology, 1964, 38(3-4), 173-182.

* Speak, M. Communication failure in questioning: Errors, misinterpretations, and personal frames of reference. Occupational Psychology, 1967, 41(4), 169-181.

Investigator Error, Interviews	4
ORA	R-M

Spector, A. J. Influences on merit ratings. Journal of Applied Psychology, 1954, 38, 393-396.

Investigator Error	13
Psych. Abst., 29, #6336	A-M

Spector, A. J. Forced-choice and projective techniques in attitude measurement (AD 134 221). Maxwell AFB, Ala.: Officer Education Research Laboratory, 1957(a).

N/A	N/A
N/A	R-NA

* Spector, A. J. The user's role in constructing a human relations test. Personnel Psychology, 1957 (b), 10, 145-156.

Attitude Measures, Questionnaire Theory and Development	8, 10, 12
Psych. Abst., 33, #2305	A-M

Spiritas, A. A., & Holmes, D. S. Effects of models on interview responses. Journal of Counseling Psychology, 1971, 18(3), 217-220.

Staats, A. W. Names as reinforcers: The social value of verbal stimuli (AD 719 414). Honolulu, Hawaii: Hawaii University, 1970. Report No. TR-9.

N/A	N/A
N/A	T-M

Stagner, R. The cross-out technique as a method in public opinion analysis. Journal of Social Psychology, 1940, 11, 79-90.

X-O Test Items	2
ORA	R-H

Stagner, R., & Osgood, C. E. Impact of war on a nationalistic frame of reference: I. Changes in general approval and qualitative patterning of certain stereotypes. Journal of Social Psychology, 1946, 24, 187-215.

Stalnaker, J. M., et al. Construction and application of psychological tests in the armed services. Review of Educational Research, 1944, 14, 102-109.

* Stangenberg, C. New definitions of scale-types. Theoria: A Swedish Journal of Philosophy, 1966, 32(1), 56-61.

Scaling	14
Psych. Abst., 40, #12708	A-H

Stanley, J. C. A comparison of verbal and pictorial self-rating scale categories. Journal of Experimental Education, 1955, 23, 239-246.

Question Stem	3g
Psych. Abst., 30, #1918	A-H

Stanley, J. C., & Wang, M. D. Differential weighting: A survey of methods and empirical studies (ERIC Document Reproduction Service, ED 030 148). Baltimore, Maryland: John Hopkins University, Center for the Study of Social Organization of Schools, 1968.

Scoring	18
ERIC Document Reproduction Service, ED 030 148	A-N

Stanton, F. Notes on the validity of mail questionnaire returns. Journal of Applied Psychology, 1939, 23, 95-104.

* Stanton, F., & Baker, K. H. Interviewing bias and the recall of incompletely learned materials. Sociometry, 1942, 5, 123-134.

Interviews, Investigator Error	12
ORA from summary in Psych. Abst., 26, #832	A-M

Stanton, H. E. Rating or inventory: A comparison of two approaches to personality measurement. Australian Psychologist, 1972, 7(1), 33-39.

Starry, A. R., & others. Stability ratings as classifiers of life history item retest reliability. Journal of Applied Psychology, 1969, 53(1), 14-18.

Stary, D. Does the modified method of forced choice eliminate the rater's bias. Revija za Psiholgiju, 1970, 1(1), 19-22.

N/A	N/A
N/A	T-H

* Staugas, L., & McQuitty, L. L. A new application of forced-choice ratings. Personnel Psychology, 1950, 3, 413-424.

Forced Choice Items, Rating Scale	2, 14
Psych. Abst., 25, #3985 (Rev. from rept.)	R-M

Stebbins, R. A. The unstructured research interview as incipient interpersonal relationship. *Sociology and Social Research*, 1972, *56*(2), 164-179.

* Steele, H. L. On the validity of projective questions. *Journal of Marketing Research*, 1964, *1*, 46-49.

Attitude Measures, Interviews, Open-ended Items, Validity, Projective Items — 1, 2, 10, 3g

ORA — R-H

Stefflre, B. The reading difficulty of interest inventories. *Occupations*, 1947, *26*, 95-96.

* Steinbock, C. B. A comparison of three item formats with respect to reliability, fakability, and acceptability to respondents. *Dissertation Abstracts International*, 1972, *33*(5-A), 2182-2183.

Rating Scales, Ranking, Check List, Respondent's Motivation — 2, 11

Dissertation Abstracts International, *33*(5-A), pp. 2182-2183 (Rev.) — A-M

Steiner, I. D. Scalogram analysis as a tool for selecting poll questions. *Public Opinion Quarterly*, 1955, *19*, 415-424.

Question Stem — 3g

Psych. Abst., *32*, #380 — A-M

Steinheiser, F. H. Individual preference scales within a multidimensional "similarities" space. *Journal of Experimental Psychology*, 1970, *86*, 325-327.

Steinmetz, H. C. Measuring ability to fake occupational interest. *Journal of Applied Psychology*, 1932, *16*.

* Stember, H., & Hyman, H. How interviewer effects operate through question form. International Journal of Opinion and Attitude Research, 1949, 3, 493-512.

Investigator Error, Response Alternative, Respondent's Motivation — 3a, 13, 11

Psych. Abst., 24, #5793 — A-H

Stember, H., & Hyman, H. Interviewer effects in the classification of responses. Public Opinion Quarterly, 1949-50, 13(4), 669-682.

Stephan, F. F., & McCarthy, P. J. Sampling opinions: An analysis of survey procedure. New York: John Wiley, 1963.

* Stephenson, W. The study of behavior. Chicago: University of Chicago Press, 1953.

Textbook, Card sorts — 17

ORA — R-M

Stern, C. G. Congruence and dissonance in the ecology of college students. Student Medicine, 1960, 8, 304-339.

Stevens, S. N., & Wonderlic, E. F. An effective revision of the rating technique. Personnel Journal, 1934, 13, 125-134.

* Stevens, S. S., On the theory of scales of measurement, Science, 1946, 103 (2684), 677-680.

Data Analysis, Instrument Format — 8, 17, 1

Abaum, et al. (Eds.) Scientific Marketing Research, New York: Scott Foresman, 1972. (Rev. from rept.) — R-N

Stevens, S. S. Mathematics, measurement, and psychophysics. In Stevens, S. S. (Ed.), *Handbook of experimental psychology*. New York: John Wiley, 1951.

Stevens, S. S. Ratio scales, partition scales and confusion scales. In Gulliksen, H., & Messick, S. J. (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960.

Stevens, S. S. *Psychophysics and social scaling*. Morristown, N. J.: General Learning, 1972.

Steward, V. *The problem of detecting "fudging" on vocational interest tests*. Los Angeles, Calif.: Personnel Reports for Sales Executives, 1947.

Stewart, L. H., & Ronning, R. R. The use of anchors in equisection scaling of interests. *California Journal of Educational Research*, 1968, *19*(3), 127-131.

Response Alternatives, Scaling, Interest Measures 3f, 14

ORA R-M

Stewart, N., & Nelson, B. *Methodological investigation of the forced-choice technique utilizing the officer description and the officer evaluation blanks*. Study No. 701: PRS-SSU, EAR-NS-e b PR 4061-04. Adjutant General's Office, 1945.

N/A N/A

N/A T-H

Stezl. I. On the reliability of difference scores from scales with partly identical items. *Zeitschrift fur experimentelle und angewandte psychologie*, 1971, *18*(1), 157-165.

Stocking, M. *Short tailored tests*. Research Bulletin 69-62 and Office of Naval Research Technical Report, Contract N-00014-69-C-0017. Princeton, N.J.: Educational Testing Service, 1969.

Stokes, S. M., & Lehman, H. C. Influence of self-interest upon questionnaire replies. *School and Society*, 1930, *32*, 435-438.

Stone, L. A. Magnitude estimation and numerical category scale evaluations of category scale adjectival stimuli on three clinical judgmental continua. *Journal of Clinical Psychology*, 1970, *26*(1), 24-27.

Stouffer, S. A. *An experimental comparison of statistical and case history methods of attitude research*. Unpublished doctoral thesis, University of Chicago, 1930.

Stouffer, S. A. Studying the attitudes of soldiers. *Proceedings of the American Philosophical Society*, 1948, *92*, 336-340.

Attitude Measures, Military Personnel 14

Psych. Abst., *23*, #3170 A-M

Stouffer, S. A., Borgatta, E. F., Hays, D. G., & Henry, A. F. A technique for improving cumulative scales. *Public Opinion Quarterly*, 1952, *16*, 273-291.

* Stouffer, S. A., et al. *Studies in social psychology in World War II. Volume 4. Measurement and prediction*. Princeton, New Jersey: Princeton University Press, 1949.

Military Personnel, Scaling, Questionnaire Theory and Development, Scoring, Textbook 2, 8, 14

ORA R-M

Stover, R. E. The measurement of change in a unidimensional attitude by Guttman scale analysis techniques. *Public Opinion Quarterly*, 1958, *22*, 116-122.

Military Personnel, Instrument Format 3g

Psych. Abst., *33*, #10143 (Rev. from rept.) R-H

* Strahan, R. Subject satisfaction with questionnaire assessment a function of response format. *Proceedings of the Annual Convention of the American Psychological Association*, 1971, *6* (Pt. 1), 125-126.

Response Alternatives, Rating Scales, True-False Items 3f, 3a

ORA R-H

Straits, B. C., Wueben, P. L., & Theophile, J. M. Influences on subjects' perceptions of experimental research situations. *Sociometry*, 1972, *35*(4), 499-518.

* Stricker, L. J. Some item characteristics that evoke acquiescent and social desirability response sets on psychological scales. *Dissertation Abstracts*, 1962, *22*(11), 4077-4078.

Attitude Measures, Personality Measures, Response Bias 12,14

Dissertation Abstracts, *22*(11), pp. 4077-4078 (Reve) A-M

* Stricker, L. J. Acquiescence and social desirability response styles, item characteristics, and conformity. *Psychological Reports*, 1963, *12*, 319-341.

Response Bias, Attitude Measures, Personality Measures, Clarity 12, 4

Psych. Abst., *38*, #4225 A-H

Stricker, L. J. *Sequential trends in response styling*. Research Memorandum 64-7. Princeton, New Jersey: Educational Testing Service, 1964.

N/A N/A

N/A T-M

Strong, E. K., Jr. Procedure for scoring an interest test. *Psychol. Clinic*, 1930, *19*, 63-72.

Stuart, A. *Basic ideas of scientific sampling*. London: Griffin, 1962.

Stycos, J. M. Interviewer effect on scale reproducibility. *American Sociological Review*, 1955, *20*(4), 443-446.

Sublette, D. J. The preparation of pencil and paper tests. *Public Personnel Review*, 1941, *2*, 1-17.

Suchman, D. I. Responses of subjects to two types of interviews. *Dissertation Abstracts*, 1967, *27*(9-B), 3297.

Interviews 2

Dissertation Abstracts, *27*(9-B), p. 3297. A-M

Suchman, E. A. The intensity component in attitude and opinion research. In Stouffer, S. A., et al. (Eds.), *Measurement and prediction*. Princeton, New Jersey: Princeton University Press, 1950.

* Suchman, E. A., & Guttman, L. A solution to the problem of question "bias". *Public Opinion Quarterly*, 1947, *11*, 445-455.

Investigator Error 13

Potter, Sharpe, Hendee and Clark, 1972 A-H

Suchman, E. A., & McCandless, B. Who answers questionnaires? *Journal of Applied Psychology*, 1940, *24*, 758-769.

Suckinan, E. A. *Scale analysis and the intensity component in attitude and opinion research*. Princeton, N. J.: Princeton University Press, 1950.

Sudman, S. On the accuracy of recording of consumer panels-Part II. *Journal of Marketing Research*, 1964, *1*(3), 69-83.

Sudman, S. New approaches to control of interviewing costs. *Journal of Marketing Research*, 1966, *3*(1), 56-61.

* Sudman, S., & Bradburn, N. M. Response effects in surveys: A review and synthesis. Chicago: Aldine Publishing Co., 1974.

Response Bias, Interviews	12
ORA	A-H

Sudman, S., & Ferber, R. A comparison of alternative procedures for collecting consumer expenditure data for frequently purchased products. Urbana, Ill.: Illinois University, Survey Research Lab., 1973. Faculty Working Paper No. 87.

* Sudman, S., Greeley, A., and Pinto, L. The effectiveness of self-administered questionnaires. Journal of Marketing Research, 1965, 2 293-297.

Interviews, Attitude Measures	1
ORA	R-H

* Summers, G. F., & Hammonds, A. D. Effect of racial characteristics of investigator on self-enumerated responses to a Negro prejudice scale. Journal of Social Forces, 1966, 44(4), 515-518.

Attitude Measures, Investigator Error	9, 12
Journal of Social Forces, 44, p. 515 (Rev. from rept.)	R-H

Sundland, D. M. The construction of Q sorts: A criticism. Psychological Review, 1962, 69, 62-64.

Survey Research Centre. Respondent understanding of questions in the survey interviews. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science, 1967 (a).

N/A	N/A
N/A	T-H

Survey Research Centre. The semantic differential scaling system in market research: I. Order effects. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science, 1967(b).

N/A	N/A
N/A	T-M

Survey Research Center. Interviewer's manual. Ann Arbor, Mich.: University of Michigan, Survey Research Center, Institute for Social Research, 1969 (a).

Survey Research Centre. The semantic differential scaling system in market research: III. Interviewer deviation from instructions. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science, 1969 (b).

N/A	N/A
N/A	T-M

Survey Research Centre. A comparison of the vertical and horizontal systems of presenting differential rating scales. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science, 1970.

Semantic Differential Items, Instrument Format, Response Alternatives	3e
ORA	R-H

* Survey Research Centre. The extent and the nature of order effects in using the semantic differential scaling technique. London: Reprint Series, Survey Research Centre, London School of Economics and Political Science, 1972.

Instrument Format, Semantic Differential Items, Question Stem	3c, 9
ORA	R-H

Sutcliffe, J. P., & Bristow, R. A. Do rank order and scale properties remain invariant under changes in the set of scaled stimuli? Australian Journal of Psychology, 1966, 18, 26-40.

N/A N/A

N/A T-H

Swineford, F. 1971 AERA Conference summaries: Innovations in measurement. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement and Evaluation, 1972. Report No. TM-R-15.

* Swordes, A. Effect of changing the number of item responses from five to four in the same test. Journal of Applied Psychology, 1952, 36, 342-343.

Multiple Choice Items, Response Alternatives, Clarity, Instrument Format, Investigator Error 2, 3g, 13

Psych. Abst., 27, #4752 A-H

* Symonds, P. M. On the loss of reliability in ratings due to coarseness of the scale. Journal of Experimental Psychology, 1924, 7, 456-461.

Data Analysis, Personality Measures, Reliability, Response Alternatives, Scaling 3a, 8

ORA R-H

* Symonds, P. M. Influence of order of presentation of items in ranking. Journal of Educational Psychology, 1936, 27, 445-449.

Attitude Measures, Instrument Format, Ranking 3b, 3c, 9

ORA R-H

Symonds, P. M. Research on the interviewing process. Journal of Educational Psychology, 1939, 20, 346-353.

Szaly, L. B., et al. Attitude measurement by free verbal association. Journal of Social Psychology, 1970, 82(1), 43-55.

* Tajfel, H. The anchoring effects of value in a scale of judgments. British Journal of Psychology, 1959, 50, 294-304.

Investigator Error, Response Bias, Raters, Response Alternatives 3f, 12

ORA R-H

Tajfel, H., Richardson, A., & Everstine, L. Individual consistencies in categorizing: A study of judgment behavior. Journal of Personality, 1964, 32, 90-108.

Tallent, N. A note on an unusually high rate of returns for a mail questionnaire. Public Opinion Quarterly, 1959, 23, 579-581.

Tamir, P. An alternative approach to the construction of multiple choice test items. Journal of Biological Education, 1971, 5(6), 305-307.

Tate, M. W., & Clelland, R. C. Nonparametric and shortcut statistics. Danville, Ill.: Interstate Printers and Publishers, 1957.

Data Analysis 18

ORA R-N

Taylor, E. K., & Hastman, R. Relation of format and administration to the characteristics of graphic rating scales. Personnel Psychology, 1956, 9(1), 181-206.

Ranking, Raters, Instrument Format, Scoring 3c, 8, 3g

ORA R-H

Taylor, E. K., & Manson, G. E. Supervised ratings - making graphic scales work. Personnel, 1951, 27, 504-514.

Taylor, E. K., et al. Rating scale content: II Effect of rating on individual scales. Personnel Psychology, 1958, 11, 519-533.

Taylor, I. A. Similarities in the structure of extreme social attitudes. Psychological Monographs, 1960, 74(2), Whole No. 489.

Attitude Measures	10
ORA	R-N

* Taylor, J. B., & Parker, H. A. Graphic ratings and attitude measurement: A comparison of research tactics. Journal of Applied Psychology, 1964, 48(1), 37-42.

Attitude Measures, Rating Scales, Reliability	2
Psych. Abst., 38, #8297	A-H

Taylor, J. C., & Bowers, D. G. The survey of organizations: Towards a machine-scored, standardized questionnaire format. Ann Arbor, Mich.: University of Michigan, Institute for Social Research, 1970. ONR Contract N00014-67-A-0181-0013.

N/A	N/A
N/A	T-M

TenHouten, W. D. Scale gradient analysis: A statistical method for constructing and evaluating Guttman Scales. Sociometry, 1969, 32(1), 80-98.

* Tenopyr, M. L. Internal consistency of ipsative scores: The "one reliable scale" phenomenon. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 3, 245-246.

Forced Choice Items, Questionnaire Theory and Development, Reliability	14
ORA	R-M

* Terris, F. Are poll questions too difficult? Public Opinion Quarterly, 1949, 13(2), 313-319.

Clarity	4
ORA	R-H

Thomas, J. A., & Sadacca, R. Impact of feedback on accuracy of confidence levels assigned by interpreters. U. S. Army BESRL Technical Research Note No. 187, 1967.

Thompson, C. A. Development of the Airman Qualifying Examination, forms D and E. United States Air Force WADC, Technical Report No. 58-94, Pt. 1, 1958.

Questionnaire Theory and Development	18
Psych. Abst., 33, #11109	A-NA

Thompson, J. W. Bi-polar and unidirectional scales. British Journal of Psychology, 1963, 54(1), 15-24.

Thorndike, R. L. Personnel selection; test and measurement techniques. New York: Wiley, 1949.

Questionnaire Theory and Development, Textbook	17, 14
Psych. Abst., 23, #5074	A-M

* Thorndike, E. L., & Lorge, R. The teacher's word book of 30,000 words. New York: Columbia University Press, 1944.

Clarity	15
ORA	R-N

* Thumin, F. J. Watch for those unseen variables. Journal of Marketing, 1962, 26, 59.

Instrument Format, Question Stem, Interviews, Respondent's Motivation	3c, 7, 9, 11, 12, 3g
ORA	R-M

Thurstone, L. L. Equally often noticed differences. *Journal of Educational Psychology*, 1927 (a), *18*, 289-293.

Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927 (b), *34*, 273-286.

Questionnaire Theory and Development 14

ORA R-N

Thurstone, L. L. The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 1927 (c), *21*, 384-400.

Thurstone, L. L. Attitudes can be measured. *American Journal of Sociology*, 1928 (a), *33*, 529-554.

Thurstone, L. L. The measurement of opinion. *Journal of Abnormal and Social Psychology*, 1928 (b), *22*, 415-420.

Questionnaire Theory and Development, Scaling 15

Journal of Abnormal and Social Psychology, *22*, p. 430 R-N

Thurstone, L. L. The indifference function. *Journal of Social Psychology*, 1931(a), *2*, 139-167.

Questionnaire Theory and Development 18

ORA R-N

Thurstone, L. L. The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, 1931 (b), *26*, 249-269.

Thurstone, L. L. *The measurement of social attitudes*. Chicago: University of Chicago Press, 1931(c).

Textbook 16

ORA T-H

* Thurstone, L. L. *The measurement of values*. Chicago: University of Chicago Press, 1959.

Scaling 17

ORA R-H

Thurstone, L. L., & Chave, E. J. *The measurement of attitudes*. Chicago: University of Chicago Press, 1929.

Tiffin, J. Merit rating: Its validity and techniques. In Dooher, M. J., & Marquis, V. (Eds.), *Rating Employee and Supervisory Performance, A Manual of Merit Rating Techniques*. New York: American Management Association, 1950.

Timothy, R. Relationships of self-ratings and external judgments to relatively objective measurements, *University of Southern California, Abstracts of Dissertations...1948*, 1948, 286-291.

Tireman, L. S., & Woods, V. E. Note on the influence on the validity of a vocabulary test of the method of indicating responses. *Journal of Educational Psychology*, 1940, *31*, 153-154.

Tittle, C. R., & Hill, R. J. Attitudes measurement and the prediciton of behavior: An evaluation of conditions and measurement techniques. *Sociometry*, 1967, *30*, 199-213.

Titus, H. E., & Hollander, E. P. The California F scale in psychological research: 1950-1955. *Psychological Bulletin*, 1957, *54*, 47-64.

Toman, W. The Multiple Attitude Test: A diagnostic device. *Journal of Abnormal and Social Psychology*, 1955, *51*, 163-170.

Questionnaire Theory and Development 18

Psych. Abst., *30*, #4604 A-N

* Toops, H. A. The factor of mechanical arrangement and typography in questionnaires. Journal of Applied Psychology, 1937, 21(2), 225-229.

Clarity, Instrument Format, Open-Ended Items	4, 8, 9, 14
Potter, Sharpe, Hendee and Clark, 1972 (Rev. from rept.)	R-N

Toops, H. A. A comparison by work-limit and time-limit of item analysis indices for practical test construction. Educational and Psychological Measurement, 1960, 20, 251-266.

Achievement Measures	18
Psych. Abst., 35, #6395	A-N

* Torgerson, W. S. Theory and methods of scaling. New York: John Wiley, 1958.

Scaling, Questionnaire Theory and Development	17
ORA	A-H

Torgerson, W. S. Scaling and test theory. Annual Review of Psychology, 1961, 12, 51-70.

Towne, D. C. Displaying semantic differential data in three-dimensional space. Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, 1971.

Traub, R. E., & Hambleton, D. K. The effect of scoring instructions and degree of speediness on the validity and reliability of multiple-choice tests. Education & Psychological Measurement, 1972, 32 (3), 737-758.

Achievement Measures	18
ERIC Document Reproduction Service, EJ 064 145	A-N

* Travers, R. M. W. A critical review of the validity and rationale of the forced-choice technique. Psychological Bulletin, 1951, 48, 62-70.

Forced-Choice Items, Rating Scales, Raters 2

Psych. Abst., 25, #664 A-M

* Trent, R. The color of the investigator as a variable in experimental research with negro subjects. The Journal of Social Psychology, 1954, 40 (2), 281-287.

Investigator Error 9, 13

ORA R-M

Tressalt, M. E., & Volkmann, J. The production of uniform opinion by non-social stimulation. Journal of Abnormal and Social Psychology, 1942, 37, 234-243.

Triandis, H. C. Attitude and attitude change. New York: John Wiley, 1971.

Trott, D. M., & Jackson, D. N. An experimental analysis of acquiescence. Journal of Experimental Research in Personality, 1967, 2, 278-288.

Trow, M. Comment on 'participant observation and interviewing-a comparison.' Human Organization, 1957, 16(3), 33-35.

* Tsudzuki, A. Shitsumonshi chōsahō ni kansuru kenkyū II; muōtō no bunseki. (Studies on the questionnaire method II; analysis of non-response.) Japanese Journal of Psychology, 1953, 24, 226-238.

Tucker, L. R. Description of the choices of most preferred among triples of stimuli by a preference space generated from choices among pairs of stimuli. Research Memorandum 58-8. Princeton, New Jersey: Educational Testing Service, 1957.

N/A N/A

N/A T-M

Tucker, L. R. Scaling and test theory. Annual Review of Psychology, 1963, 14, 351-364.

Literature Review	16
ORA	R-M

* Tuckman, J., & Lorge, I. The effect of changed directions on the attitudes about old people and the older worker. Educational and Psychological Measurement, 1953, 13, 574-595.

True-False Items, Response Alternatives, Scoring, Rating Scales, Respondent's Motivation	2, 3a, 11
ORA	R-H

Tupes, E. C., & Christal, R. E. Stability of personality trait rating factors obtained under diverse conditions. USAF WADS Technical Note No. 58-61, 1958.

* Turgut, M. F. A comparison between two forced-choice personality test formats. Dissertation Abstracts, 1963, 24(5), 2118-2119.

Card Sorts, Paired Comparison Items, Personality Measures, Respondent's Motivation	3g, 2, 11
Dissertation Abstracts, 24(5), pp. 2118-2119 (Rev.)	A-H

Turner, C. B., & Fiske, D. W. Item quality and appropriateness of response processes. Educational and Psychological Measurement, 1968, 28, 297-315.

Clarity, Reliability	3g, 4
Educational and Psychological Measurement, 28, p. 314	R-NA

Tversky, A. Intrasitivity of preferences. Psychological Review, 1969, 76, 31-48.

Tversky, A. On the optimal number of alternatives at a choice point. Journal of Mathematical Psychology, 1964, 1, 386-391.

Multiple Choice Items, Response Alternatives 3a

ORA R-M

Twedt, D. W. Consumer psychology. Annual Review of Psychology, 1965, 16, 265-294.

Tyler, L. E. Tests and measurements. (2nd ed.) Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Udell, J. Can attitude measurement predict consumer behavior? Journal of Marketing, 1965, 29, 46-50.

Attitude Measures, Scaling, Validity 14

ORA R-M

Uhrbrock, R. S. 2000 scaled items. Personnel Psychology, 1961, 14(2), 375-420.

Rating Scales, Card Sorts 14

Psych. Abst., 37, #3919 R-M

United States Government Printing Office. Health survey procedure-concepts, questionnaire development and definitions in the health interview survey. Washington, D. C.: U.S. Government Printing Office, Health Statistics, Series 1, No. 2, 1964.

Interviews 18

ORA R-NA

* Upshaw, H. S. Own attitude as an anchor in equal-appearing intervals. *Journal of Abnormal and Social Psychology*, 1962, *64*, 85-96.

Response Bias, Response Alternatives	12, 3f
ORA	R-H

Upshaw, H. S. The effect of variable perspectives on judgments of opinion statements for Thurstone scales: Equal appearing intervals. *Journal of Personality and Social Psychology*, 1965, *2*, 60-69.

Data Analysis, Scaling, Response Bias	12, 14
Journal of Personality and Social Psychology, *2*, p. 60	R-N

Upshaw, H. S., Ostrow, T. M., & Ward, C. D. Content versus self-rating in attitude research. *Journal of Experimental Social Psychology*, 1970, *6* (3), 272-279.

Uriel, G. F. Scale and intensity analysis in opinion research. *International Journal of Opinion and Attitude Research*, 1950, *4*, 192-208.

Urry, V. W., & Nicewander, W. A. Factor analysis of the commander's evaluation report. *U.S. Army Enlisted Evaluation Center Technical Report No. 40*, 1966.

* U. S. Army Test and Evaluation Command. *Development of a guide and checklist for Human Factors Evaluation of Army equipment and systems*. U. S. Army Test and Evaluation Command (TECOM), 1973. Contract DAAD05-73-C-0388.

Adjectives	6
ORA	R-NA

U. S. Dept. Army, AGC, PRB. A study of officer rating methodology: IV. Effect of forced choice items on validity of rating scales. Personnel Research Branch Report No. 903, 1952 (a).

Forced Choice Items, Military Personnel,
Raters, Rating Scales 1, 2

Psych. Abst., 28, #8182 A-M

* U. S. Dept. Army, AGO, PRB. A study of officer rating methodology. VIII. Validity of two types of rating techniques: Forced-choice items and rating scales. Personnel Res. Br. Rep., No. 907. Washington: American Documentation Institute, 1952 (b).

Check List, Forced Choice Items, Military Personnel, 1, 2
Raters, Rating Scales

Psych. Abst., 28, #8186 A-M

Van Der Veen, F., Howard, K. I., & Austria, A. M. Stability and equivalence scores based on three different response formats. Proceedings of the 78th Annual Convention of the American Psychological Association, 1970, 5, 99-100 (Summary)

Card Sorts, Multiple Choice Items, True-False Items, 2, 8, 12
Response Alternatives, Reliability, Response Bias

ORA R-H

van Naerssen, R. F. Two-choice items in study-tests. Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 1970, 25(6), 393-403.

Vaughn, C. L. A scale for assessing socio-economic status in survey research. Public Opinion Quarterly, 1958, 22, 19-34.

Vaughn, G. M. Eliminating a non-scale type from a social distance scale. Psychological Reports, 1962, 11(3), 912.

Attitude Measures, Clarity 3g

ORA R-H

Veldman, D. J., & Parker, G. V. Adjective rating scales for self description. Multivariate Behavioral Research, 1970, 5(3), 295-302.

Adjective Rating Scales 18

Multivariate Behavioral Research, 5, p. 295 A-NA

Vernon, P. E. Questionnaires, attitude tests, and rating scales. In F. S. Bartlett, et al. (Eds.) The study of society. London: Kegan Paul, Trench, Trubner, 1939 and 1946.

Questionnaire Theory and Development 14

ORA R-N

Vernon, P. E. Personality tests and assessments. London: Methuen, 1962.

Vernon, P. E. Effects of administration and scoring on divergent thinking tests. British Journal of Education Psychology, 1971, 41(3), 245-257.

Vicary, J. M. The circular test of bias in personal interview surveys. Public Opinion Quarterly, 1955, 19, 215-218.

Vidich, A. J., & Shapiro, G. A comparison of participant observation and survey data. American Sociological Review, 1955, 20, 28-33.

Vinacke, W. E. Two tests to measure exploitative and accommodative strategy. Buffalo, N.Y.: State University of New York, 1964, Technical Report No. 7.

Respondent's Motivation 18

ORA A-NA

Vincent, N. L. The development of a new attitude measurement technique. Dissertation Abstracts, 1961, 21, 1997.

Voas, R. B. A procedure for reducing the effects of slanting questionnaire responses toward social acceptability. *Educational and Psychological Measurement*, 1958 (a), *18*, 337-345.

Voas, R. B. *Relationships among three types of response sets*. USN School of Aviation Medical Projects in Research, Subtask 1, Prj. No. NM011, 1958 (b). Report No. 15.

Voas, R. B., Blari, J. T., & Ambler, R. K. *Validity of personality inventories in the Naval aviation selection program*. USN School of Aviation, Subtask 1, Prj. No. NM 16 01 11, 1957. Report No. 13.

Military Personnel, Personality Measures 18

Psych. Abst., *33*, #853 A-NA

Volkmann, J. The method of single stimuli. *American Journal of Psychology*, 1932, *44*, 808-809.

Volkmann, J. The relation of the time of judgment to the certainty of judgment. *Psychological Bulletin*, 1934, *31*, 672-673.

* Volkmann, J. The anchoring of absolute scales. *Psychological Bulletin*, 1936, *33*, 742-743.

Instrument Format, Investigator Error 3f

ORA R-H

Volkmann, J. The natural number of categories in absolute judgment. *Psychological Bulletin*, 1937, *34*, 543-544.

Response Alternatives 3a

ORA R-H

Volkmann, J. The compression of an absolute scale. *Psychological Bulletin*, 1938, *35*, 676.

Volkmann, J. *Lectures in the psychology of judgment*. Columbia University, 1942. (Unpublished.)

N/A N/A

N/A T-M

Volkmann, J. Scales of judgment and their implications for social psychology. In Rohrer, J. H., & Sherif, M. *Social psychology at the crossroads; the University of Oklahoma lectures in social psychology*. New York: Harper, 1951. Pp. 273-298.

Rating Scales, Textbook 15

Psych. Abst., *26*, #836 (Rev. from rept.) R-N

Volkmann, J., Hunt, W. A., McGourty, M. Variability of judgment as a function of stimulus-density. *American Journal of Psychology*, 1940, *53*, 277-284.

Votaw, D. F., & Danforth, L. The effect of method of response upon the validity of multiple-choice tests. *Journal of Educational Psychology*, 1939, *30*, 624-627.

Wagner, I. F. Articulate and inarticulate replies to questionnaires. *Journal of Applied Psychology*, 1939, *23*(1), 104-115.

Wagner, R. F. A group situation compared with individual interviews for securing personnel information. *Personnel Psychology*, 1948, *1*, 93-107.

Wagner, R. K. *An investigation of government employee perceptions of their organizational climate*. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, 1971. Report No. GSM SM 71-13.

Interviews, Military Personnel, Preference Measures 18

ORA A-NA

Wahler, H. J. Response styles in clinical and nonclinical groups. Ju Journal of Consulting Psychology, 1961, 25, 533-539.

Wakefield, J. A. A reply to Dr. Adkins. Public Personnel Review, 1958(b) 18, 298-300.

Wakefield, J. A. Does the fifth choice strengthen a test item? Public Opinion Quarterly, 1958(a), 19, 45-48.

Achievement Measures 18

Psych. Abst., 33, #7029 A-N

Waksberg, J. Conditioning effects from repeated household interviews. Journal of Marketing, 1964, 28(2), 51-56.

Walker, D. A. Answer-pattern and score-scatter in tests and examinations. British Journal of Psychology, 1940, 30, 248-260.

Walker, H. W. Certain mathematical questions suggested by true-false test. Amer. Math. Mo., 1927, 34, 504-515.

Walker, K. P. Examining personnel information items of a questionnaire study. Journal of Educational Research, 1937, 31, 281-282.

Wallace, D. Mail questionnaires can produce good samples of homogeneous groups. Journal of Marketing, 1947, XII(1), 53-60.

Walsh, J. A., Jr. A factorial study of a large sample of response set and attitude scales. Dissertation Abstracts, 1964, 24(11), 4824-4825.

Walsh, J. A. Prediction of anchor effects on personality items from rating dispersions. Educational and Psychological Measurement, 1968, 28(2), 317-325.

Personality Measures, Response Alternatives 3f, 14

ORA R-H

* Walsh, W. B. Validity of self-report. Journal of Consulting Psychology, 1967, 14(1), 18-23.

Interviews, Closed-Ended Items, Respondent's Motivation	1
ORA	R-N

Walsh, W. B. Validity of self-report: Another look. Journal of Counseling Psychology, 1968, 15(2), 180-186.

Walsh, W. B. Self-report under socially undesirable and distortion conditions. Journal of Counseling Psychology, 1969, 16(6), 569-574.

Walters, J. H. Structured or unstructured techniques? Journal of Marketing, 1961, 25, 58-62.

Open-Ended Items	1
ORA	R-M

Wang, C. K. A. Suggested criteria for writing attitude statements. Journal of Social Psychology, 1932, 3(3), 367-373.

Ward, C. D. Ego-involvement and the absolute judgment of attitude statements. Paper to American Psychological Association, Annual Meetings, Philadelphia, 1963. (mimeographed)

N/A	N/A
N/A	T-H

Ward, C. D. Multiple-choice question writing: Research participation and exam performance. Journal of College Science Teaching, 1973, 3(1), 77-78.

* Ward, C. D. Issue saliency and the correspondence between measures of attitude. College Park, Maryland: Maryland University, Dept. of Psychology, 1969. Report No. TR-15.

Scaling, Semantic Differential Items,
Attitude Measures 2, 10

DDC, # AD 695 000 A-H

Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 1965, 60(309), 63-69.

* Warren, G. S. Item content and format as evidence for response biases. Dissertation Abstracts International, 1972, 33(6-A), 2778.

Investigator Error, Response Bias 10, 12

ORA R-H

Warren, R. D., & others. Moderator effects on attitude scale construction. Home Economics Research Journal, 1973, 1(4) 259-268.

Waters, C., & Waters, L. K. Effect on number of alternatives and scoring instructions on examinees' reactions to multiple-choice tests. Psychological Reports, 1971(a), 29, 1229-1330.

Waters, C., & Waters, L. K. Validity and likability ratings for three scoring instructions for a multiple-choice vocabulary test. Educational and Psychological Measurement, 1971 (b), 31(4), 935-938.

* Waters, L. K. Effects of instructions and item tone to forced-choice pairs. Personnel Psychology, 1966, 19(1), 45-53.

Rating Scales, Forced-Choice Items,
Respondent's Motivation 2, 7, 11, 3g

ORA R-H

* Waters, L. K., & Wherry, R. J., Jr. Evaluation of two forced-choice response formats. Personnel Psychology, 1961(a), 14(4), 285-289.

Forced-Choice Items, Rating Scales, Instrument Format, Respondent's Motivation 2, 11, 3g

Psych. Abst., 37, #5734 A-H

* Waters, L. K., & Wherry, R. J. A note on the stability of the preference index in forced-choice blocks. USN School of Aviation Medical Research Report No. 3, 1961(b).

Check List, Forced-Choice Items 2

Psych. Abst., 38, #4285 A-M

Waters, L. K., & Wherry, R. J., Jr. A note on alternative methods of scoring a forced-choice form. Personnel Psychology, 1962(a), 15(3), 315-317.

Scoring Validity, Forced-Choice Items 8

Psych. Abst., 37, #7265 A-M

Waters, L. K., & Wherry, R. L., Jr. The preference index and responses to forced-choice pairs. Personnel Psychology, 1962(b), 15(2), 99-102.

* Watson, J. J. Improving the response rate in mail research. Journal of Advertising Research, 1965, 5(2), 48-50.

Respondent's Motivation 11

ORA K-M

Weaver, H. G. Consumer questionnaire technique. American Marketing Journal, 1934, 1(3), 115-118.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally, 1966.

Webb, J. T. Subject speech rates as a function of interviewer behavior. Language and Speech, 1969, 12(1), 54-67.

Webb, S. C. Scaling of attitudes by the method of equal-appearing intervals: A review. Journal of Social Psychology, 1955, 42, 215-239.

Webb, S. C., & Chueh, J. C. The effect of role taking on the judgment of attitude items. Paper to American Psychological Association, Annual Meetings, St. Louis, 1962. (mimeographed)

Webb, W. B. Self-evaluation compared with group evaluations. Journal of Consulting Psychology, 1952, 16, 305-307.

Rating Scales, Raters 12

Psych. Abst., 27, #4236 A-N

Webb, W. B. A procedure for obtaining self-ratings and group ratings. Journal of Consulting Psychology, 1956, 20, 233-236.

Webster, H. Correcting personality scales for response sets or suppression effects. Psychological Bulletin, 1958, 55, 62-64.

* Webster, H. The meaning of response set in personality inventories. American Psychologist, 1960, 15(7), 431. (Abstract.)

Personality Measures, Response Bias 12

American Psychologist, 15, p. 431 A-M

Webster, H. Acquiescence, social desirability and inhibition reflected by "response set" scales. Psychological Reports, 1962, 10, 789-790.

* Wedell, C., & Smith, K. U. Constancy of interview methods in appraisal of attitudes. Journal of Applied Psychology, 1951, 35, 392-396.

Interviews, Investigator Error, Rating Scales — 1, 2, 7, 13, 12

Psych. Abst., 26, #6242 (Rev. from rept.) — R-H

Weick, K. E. Systematic observational methods. In Lindzey, G., & Aronson, E. (Eds.), The handbook of social psychology. Volume 2. (2nd edition.) Reading, Mass.: Addison-Wesley, 1968. Chapter 13.

Weidemann, C. C. The omission as a specific determiner in the true-false examination. Journal of Educational Psychology, 1931, 22, 435-439.

Weidemann, C. C., & Newens, L. F. Does the compare and contrast essay test measure the same mental functions as the true-false test? Journal of General Psychology, 1933 (a), 9, 430-449.

Weidemann, C. C., & Newens, L. F. The effect of directions preceding true-false and indeterminate statement examinations upon distribution of test scores. Journal of Educational Psychology, 1933 (b), 24, 79-106.

Weiner, M., & Tobias, S. Chance factors in the interpretation of group administered multiple-choice tests. Personnel and Guidance, 1963, 41(5),435-437.

Data Analysis, Scoring, Investigator Error — 8

Psych. Abst., 39, #1771 — A-N

Weinland, J. D. Better words on rating scales. Personnel Journal, 1946, 25, 131-134.

Response Alternatives, Rating Scales — 3f

ORA — R-N

Weinstein, A. G. Predicting behavior from attitudes. Public Opinion Quarterly, 1972, 36(3), 355-360.

Weiss, D. J. Averaging: An empirical validity criterion for magnitude estimation. _Perception & Psychophysics_, 1972, _12_(5), 385-388.

Weiss, D. J., & Dawis, R. W. An objective validation of factual interview data. _Journal of Applied Psychology_, 1960, _44_, 381-385.

Weiss, D. J., Dawis, R. W., England, G. W., & Lofquist, L. H. Validity of work histories obtained by interview. _Minnesota Studies in Vocational Rehabilitation_, 1961, _12_, 1-3.

Weiss, W. The effects on opinions of a change in scale judgments. _Journal of Abnormal and Social Psychology_, 1959, _58_, 329-334.

* Weiss, W. Effects of an extreme anchor on scale judgments and attitude. _Psychological Reports_, 1961, _8_, 377-382.

Response Alternatives, Rating Scales 3f

Psychological Reports, _36_, #2GD FFW A-H

Weiss, W. Effects of response scales on the expected distribution of judgments of social stimuli. _Psychological Reports_, 1963(a), _13_(4), 411-414.

Scaling, Response Alternatives 2, 3b

Psych. Abst., _38_, #7079 A-M

* Weiss, W. Effects of unbalanced response scales on judgments of social stimuli. _Psychological Reports_, 1963 (b), _12_(2), 403-414.

Scaling Response Alternatives 2, 3f, 3a, 14

Psych. Abst., _38_, #4156 A-H

Weiss, W. Scale judgments of triplets of opinion statements. Journal of Abnormal and Social Psychology, 1963 (c), 66, 471-479.

Investigator Error, Response Alternatives	3g, 3a
Journal of Abnormal and Social Psychology, 66, p. 471	R-NA

* Weitman, M. Forms of failure to respond and varieties of authoritarianism. Journal of Personality, 1964, 32(1), 109-118.

Attitude Measures, Open-Ended Items, Respondent's Motivation	10
Psych. Abst., 39, #5164	A-M

* Weitz, J. Verbal and pictorial questionnaires in market research. Journal of Applied Psychology, 1950, 34, 363-366.

Clarity, Preference Measures	2, 3g, 4
Psych. Abst., 26, #588 (Rev. from rept.)	R-H

* Weitz, J., & Nuckols, R. C. The validity of direct and indirect questions in measuring job satisfaction. Personnel Psychology, 1953, 6, 487-494.

Question Stem	3g
Psych. Abst., 28, #8223	A-M

Weitz, S. Attitude, voice, and behavior: A repressed affect model of interracial interaction. Journal of Personality and Social Psychology, 1972, 24(1), 14-21.

Weitzman, R. A. Pattern scoring of short predictors. Proceedings of the 81st Annual Convention of the American Psychological Association, Montreal, Canada, 1973, 8, 27-28.

N/A	N/A
N/A	T-M

* Wells, W. D. The influence of yea saying response style. *Journal of Advertising Research*, 1961, *1*, 1-12.

Response Bias, Literature Review 12

ORA R-H

* Wells, W. D. How chronic overclaimers distort survey findings. *Journal of Advertising Research*, 1963, *3*(2), 8-18.

Personality Measures, Response Bias, Literature Review 12, 16, 9

ORA R-H

Wells, W. D., & Dames, J. Hidden errors in survey data. *Journal of Marketing*, 1962, *26*(4), 50-54.

* Wembridge, E. R., & Means, E. R. Obscurities in voting upon measures due to double-negative. *Journal of Applied Psychology*, 1918, *2*, 156-163.

Question Stem 3g

Bauman, Rogers, and Weiss, 1971 A-M

Weschler, I. R., & Bernberg, R. E. Indirect methods of attitude measurement. *International Journal of Opinion and Attitude Research*, 1950, *4*, 209-228.

Wesman, A. G. Some effects of speed in test use. *Educational and Psychological Measurement*, 1960, *20*, 267-274.

Wesman, A. G. Active versus blank responses to multiple-choice items. Journal of Educational Psychology, 1947, 38, 89-95.

Check List, Multiple Choice Items 3g, 4, 3a

ORA R-H

Westfall, R. L., Boyd, H. W., & Campbell, D. T. The use of structured techniques in motivation research. Journal of Marketing, 1957, 21, 134-139.

Forced-Choice Items, Projective Items 2

ORA R-M

Wever, E. G., & Zener, K. E. The method of absolute judgments in psychophysics. Psychological Review, 1928, 35, 466-493.

Wheatley, B. C., & Cash, W. B. Employee survey: correcting its basic weakness (questionnaire technique). Personnel Journal, 1973, 52, 456-459.

* Wheatley, J. J. Self-administered written questionnaires or telephone interviews? Journal of Marketing Research, 1973, 9, 94-95.

Interviews, Semantic Differential Items 1, 2, 3g

ORA R-N

Wheeler, R. W. A study of the relationship between selected interviewer variables and the interpretation of interview information. Dissertation Abstracts International, 1969, 30 (1-B), 425.

Wherry, R. J. Orders for the presentation of pairs in the method of paired comparisons. Journal of Experimental Psychology, 1938, 23, 651-660.

N/A N/A

N/A T-H

Wherry, R. J. *Control of bias in rating: Factor analysis of rating item indices*. AGO, Personnel Research Section, 1951. Report No. 915.

Wherry, R. J., Jr. *A test of new rationale and methodology for the forced-choice technique* (AD 237 782). Pensacola, Florida: Naval School of Aviation Medicine, 1960. Project MR005 13 5001.

N/A N/A

N/A R-NA

Wherry, R. J., & Fryer, D. H. Buddy ratings: Popularity contest or leadership criteria? In Monroe, J. L. (Ed.), *The sociometry reader*. Glencoe, Illinois: The Free Press, 1960.

Whipple, J. W. A study of the extent to which positive or negative phrasing affects answers in a true-false test. *Journal of Educational Research*, 1957, *51*, 59-63.

Question Stem, Investigator Error 13

Psych. Abst., *33*, #2001 A-N

Whitlock, G. H. *Validation of morale and attitude scales* (AD 242 359). Knoxville, Tennessee: University of Tennessee, 1960. Technical Report No. 60-76.

Whitman, J. R., & Schwartz, A. N. The relationship between two measures of the tendency to give socially desirable responses. *Journal of Projective Techniques*, 1967, *31*(5), 72-75.

Whitney, D. R., & Feldt, L. S. Analyzing questionnaire results: Multiple tests of hypotheses and multivariate hypotheses. *Educational and Psychological Measurement*, 1973, *33*(2), 365-380.

Whyte, W. On asking indirect questions. Human Organization, 1957, 15(4), 21-23.

Wick, J. W. Similar response analysis. Educational and Psychological Measurement, 1970, 30(1), 95-110.

Wicker, M. P. A comparison of attitude scale values yielded by scales of differing lengths. University of North Carolina...Research in Progress. University of North Carolina Rec., 1951, 60(492), 252. (Abstract of Master's thesis.)

* Wickes, T. A., Jr. Examiner influence in a testing situation. Journal of Consulting Psychology, 1956, 20, 23-26.

Investigator Error, Respondent's Motivation 10, 11

Psych. Abst., 31, #3090 A-H

Wiener, G. The effect of distrust on some aspects of intelligence text behavior. Journal of Consulting Psychology, 1957, 21, 127-130.

Wiener, M., Carpenter, J. T., & Carpenter, B. Some determinants of conformity behavior. Journal of Social Psychology, 1957, 45, 289-297.

Wiggins, J. S. Interrelations among MMPI measures of dissumulation under standard and social desirability instructions. Journal of Consulting Psychology, 1959, 23, 319-427.

Wiggins, J. S. Convergences among stylistic response measures from objective personality tests. Educational and Psychological Measurement, 1964, 24, 551-562.

Wiggins, N. Individual viewpoints of social desirability. Psychological Bulletin, 1966, 66, 68-77.

Wight, A. R. A study of rater and ratee characteristics. *Dissertation Abstracts International*, 1969, *30*(5-B), 2451.

Personality Measures	18
ORA	A-NA

Wilbourn, J. M., & Quinn, N. *Feasibility of using special measures in the classification and assignment of lower mental ability airmen*. Brooks AFB, Texas: Air Force Human Resources Laboratory, 1973.

Achievement Measures	18
NTIS	A-N

Wilbur, P. H. Positional response set among high school students on multiple-choice tests. *Journal of Educational Measurement*, 1970, *7*, 161-164.

Wilcove, G. L. *Enlisted men's and officer's opinions of recent policy changes implemented through Z-grams*. Washington, D. C.: Naval Personnel Research and Development Laboratory, 1971. Report No. WSR-72-5.

Military Personnel	18
ORA	A-NA

Wilde, G. J. S., & DeWitt, O.E. Self-report and error choice: Inter-individual differences in the operation of the error-choice principle and their validity in personality questionnaire tests. *British Journal of Psychology*, 1970, *61*(2), 219-228.

Wiley, C. F. The three-decision multiple-choice test: A method of increasing the sensitivity of the multiple choice item. *Psychological Rep*., 1960, *7*, 475-477.

Wiley, L., Harber, H. B., & Giorgia, M. J. Evidence for a generalized rating tendency. Engineering and Industrial Psychology, 1959 (a), 1, 55-61.

Raters, Investigator Error 13

Psych. Abst., 35, #5394 (Rev.) A-N

Wiley, L., Harber, H. B., & Giorgia, J. M. Rater tendencies in estimating qualifications required by Air Force tasks. USAF WADC Personnel Laboratory, 1959(b). Technical Note No. 59-195.

N/A N/A

N/A T-H

Wiley, L., & Jenkins, W. S. Method for measuring bias in raters who estimate job qualifications. Journal of Industrial Psychology, 1963, 1(1), 16-22.

Raters, Investigator Error 12

Psych. Abst., 38, #6735 A-N

Wiley, L. N., & Trimble, O. C. The ordinary objective test as a possible criterion of certain personality traits. School and Society, 1936, 43, 446-448.

Wilkie, W. L., & Pessemier, E. A. Issues in marketing's use of multi-attribute attitude models. Journal of Marketing Research, 1973, 10, 428.

Wilkins, L. A. Suggestions as to the formulation of questions in standardized examinations in modern languages. New York Bulletin High Points, 1923, 27-36.

* Wilkins, L. T. Incentives and the young male worker in England: with some notes on ranking methodology. International Journal of Opinion and Attitude Research, 1950, 4, 541-562.

Paired Comparison Items, Ranking, Preference Measures, Military Personnel 2, 12

Psych. Abst., 26, #570 (Rev. from rept.) R-H

Will, R. T., & Hasty, R. W. Attitude measurement under conditions of multiple stimuli. *Journal of Marketing Research*, 1971, *35*, 66-70.

Preference Measures, Rating Scales 15

ORA R-NA

Willcock, H. D. Mass observation. *American Journal of Sociology*, 1943, *48*, 445-456.

Willey, C. F. The three-decision multiple-choice test: A method of increasing the sensitivity of the multiple-choice item. *Psychological Reports*, 1960, *7*, 475-477.

Williams, J. A., Jr. Interviewer-respondent interaction: A study of bias in the information interview. *Sociometry*, 1964, *27*(3), 338-352.

Williams, J. A., Jr. Interviewer role performance-A further note on bias in the information interview. In Forcese, D. P., & Richer, S. (Eds.), *Stages of social research: Contemporary perspectives*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

* Williams, W. H. The systematic bias effects of incomplete responses. *American Statistical Association, Proceedings of the Social Statistics Section*, Pittsburgh, Penn., 1968, 11th annual edition, 1968, 308-312.

Data Analysis 1

ORA R-N

Williams, W. S. A study of the use of the semantic differential by fifth grade children from different socioeconomic groups. *Journal of Psychology*, 1972, *81*(2), 343-50.

Willingham, W. W., & Jones, M. B. On the identification of halo through analysis of variance. *Educational and Psychological Measurement*, 1958, *18*, 403-407.

Willis, R. H. Manipulation of item marginal frequencies by means of multi-response items. Psychological Review, 1960, 67, 32-50.

Attitude Measures, Response Alternatives, Multiple-Choice Items, Scoring, Rating Scale	8, 2
Psych. Abst., 35, #739 (Rev. from rept.)	R-H

Willis, R. H. The phenomenology of shifting agreement and disagreement in dyads. Journal of Personality, 1965, 33, 188-199.

* Winters, S., & Bartlett, C. J. Instructional and response style factors with forced choice response. Research Reports, Sinai Hospital of Baltimore, 1966, 1, 10-16.

Response Bias, Forced-Choice Items, Investigator Error	12, 7
Psych. Abst., 40, #12852	A-H

* Winthrop, H. Reliability of preference ratings as a function of cardinal value and natural order. Psychological Reports, 1958, 4, 62.

Scaling, Reliability, Instrument Format	3b
Psychological Reports, 4, p. 62 (rev. from rept.)	R-H

* Wiseman, F. Methodological bias in public opinion surveys. Public Opinion Quarterly, 1972, 36(1), 105-108.

Interviews, Response Bias, Attitude Measures, Respondent's Motivation	1, 12, 11
ORA	R-H

Witryol, S. L., & Fischer, W. F. Scaling children's incentives by the method of paried comparisons. Psychological Reports, 1960, 7, 471-474.

* Witroyl, S. L. Scaling procedures based on the method of paired-comparisons. Journal of Applied Psychology, 1954, 38, 31-37.

Scaling	14
Psych. Abst., 29, #125	A-M

* Witroyl, S. L., & Thompson, G. C. An experimental comparison of the stability of social acceptability scores obtained with the partial-rank-order and the paired-comparison scales. Journal of Educational Psychology, 1953, 44, 20-30.

Response Alternatives, Paired Comparison Items, Reliability	2, 8
Psych. Abst., 28, #735 (Rev. from rept.)	R-H

Wofford, J. C., & Willoughby, T. L. The effects of test construction variables upon test reliability and validity. California Journal of Educational Research, 1969, 20(3), 96-106.

* Wolfe, D. F. A new questionnaire design. Journal of Marketing, 1956, 20, 186-190.

Instrument Format	15, 17
ORA	R-N

Wollack, S., Witjing, J. P., Goodale, J. G., & Smith, P. C. Weighting agreement responses by item scale values. Journal of Applied Psychology, 1970, 54(2), 174-175.

Rating Scales, Attitude Measures	2
ORA	R-H

Womer, F. B. The valuation of item selection techniques appropriate to a new response method for multiple-choice type test items. Dissertation Abstracts, 1957, 17, 98.

Wonderlic, E. F. We survey attitudes annually by mail. Personnel Journal, 1953, 32, 91-93.

Wood, D. A. Test construction, development, and interpretation of achieving tests. Columbus, Ohio: Charles E. Merrill, 1961.

Woodrow, D. Multichoice test marking scheme--Eval. Australian Science Teachers Journal, 1973, 19(2), 57-62.

Achievement Measures, Scoring 18

ERIC Document Reproduction Service, EJ 083 425 A-N

Woodson, E. J. Human factors engineering. New York: McGraw-Hill, 1964.

Woodward, J. L. Making government opinion research bear upon operation. American Sociological Review, 1944, 9, 670-677.

* Worthy, M. Note on scoring midpoint responses in extreme response-style scores. Psychological Reports, 1969, 24(1), 189-190.

Response Bias, Data Analysis, Rating Scales 8, 12

ORA R-M

* Wright, C. E. A factor dimension comparison of normative and ipsative measurements. Educational and Psychological Measurement, 1961, 21, 433-444.

Data Analysis, Scoring, Personality Measures 8

Psych. Abst., 36, #2HE33W A-M

Wright, O. R., Jr. Summary of research on the selection interview since 1940. Personnel Psychology, 1969, 22(4), 391-413.

Wrightsman, L. Characteristics of positively scored and negatively scored items from attitude scales. *Psychological Reports*, 1965, *17*, 898.

Wrigley, C., & Neuhaus, J. O. The matching of two sets of factors. *Contract Memorandum Report, A-32*. Urbana, Ill.: University of Illinois, 1955.

N/A	N/A
N/A	T-M

Wyatt, D. F., & Campbell, D. T. A study of interviewer bias as related to interviewers' expectations and own opinions. *International Journal of Opinion and Attitude Research*, 1950, *4*, 77-83.

Wylie, A. T. To what extent may we rely upon the answers to a school questionnaire? *Journal of Educational Method*, 1927, *6*, 252-257.

Yadoff, B. *An attempt to change word meaning and a personality test score through semantic generalization.* (Doctoral thesis, University of Pittsburgh) Pittsburgh, Pa. (DA 19:2157)

Yates, F. *Sampling methods for censuses and surveys*. (3rd ed.) New York: Hafner, 1960.

Yoell, W. A. How the depth interview reveals attitudes toward new products. *Printers' Ink*, 1947, *218*(6), 50-52.

Irrelevant	18
ORA	R-NA

* York, C. M. Stability of Thurstone scale values after 35 years. *Perceptual and Motor Skills*, 1966, *23*(2), 628.

Scaling, Rating Scales 2, 8

ORA R-M

Young, P. V. Schedule and the questionnaire as aids in field exploration. In *Scientific social surveys and research*, (P.V.Y.), Chapter 7. New York: Prentice Hall, 1939. Pp. 138-171.

Young, P. V. The validity of schedules and questionnaires. *Journal of Educational Sociology*, 1940, *14*, 22-26.

Young, P. V. The questionnaire and other reporting forms as aids in field exploration. In Capt., K. G. (Ed.), *Scientific social surveys and research*. (3rd ed.) Englewood Cliffs, New Jersey: Prentice Hall, 1956. Chapter VIII, pp. 176-204.

Questionnaire Theory and Development 17

ORA R-N

Young, P. V. *Scientific social surveys and research: An introduction to the background, content, methods, principles, and analysis of social studies*. (2nd ed.) Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Questionnaire Theory and Development, Textbook 17

ORA R-N

Zaccaria, M. A., Tupes, E. C., & Lawrence, H. G. *Development and characteristics of the USAF officer activity inventory*. *United States Air Force* Personnel Training, Research Center, Research Report No. 57-15, 1957.

Military Personnel, Interest Measures, Reliability 14, 18

Psych. Abst., *33*, #9202 A-NA

* Zajonc, R. B., & Nienwenhyse, B. Relationship between word frequency and recognition-perceptual process or response bias? Journal of Experimental Psychology, 1964, 67(3), 276-285.

Adjectives, Response Bias, Questionnaire Theory and Development 12, 6

ORA R-M

* Zavala, A. Development of the forced-choice rating scale technique. Psychological Bulletin, 1965, 63(2), 117-124.

Forced-Choice Items, Rating Scale, Response Bias, Respondent's Motivation 2, 12, 11

Psych. Abst., 39, #6188 A-H

* Zavalloni, M. & Cook, S. W. Influence of judge's attitudes on ratings of favorableness of statements about a social group. Journal of Personality and Social Psychology, 1965, 1(1), 43-54.

Attitude Measures, Raters 10, 13

Journal of Personality and Social Psychology, 1, p. 43 (Rev. from rept.) R-H

* Ziller, R. C., Long, B. H. Some correlates of the don't know response in opinion questionnaires. Journal of Social Psychology, 1965, 67, 139-147.

Attitude Measures, Response Bias, Response Alternatives 8, 12

Journal of Social Psychology, 67, pp. 139-147 (Rev. from rept.) R-H

* Zimbardo, P. G. Verbal ambiguity and judgmental distortion. Psychological Reports, 1960, 6, 57-58.

Raters 10

ORA R-H

Zimmer, C. E. *Chance distribution of inconsistent response patterns in paired comparison and multiple ranking designs*. USAF PRL Technical Documentation Report No. 63-1, 1963.

Scaling, Paired Comparison Items, Ranking 8, 2

Psych. Abst., *37*, #7396 A-M

* Zinnes, J. L. Scaling. *Annual Review of Psychology*, 1969, *20*, 447-479.

Literature Review, Scaling, Data Analysis 16

ORA R-M

Zubin, J. The chance element in matching tests. *Journal of Educational Psychology*, 1933, *24*, 674-681.

Zubin, J. Nomograph. *Journal of the American Statistical Association*, 1939, *34*, 539-544.

Data Analysis 18

ORA R-N

* Zuckerman, J. V. Interest item response arrangement as it affects discrimination between professional groups. *Journal of Applied Psychology*, 1952, *36*, 79-85.

Forced-Choice Items, Preference Measures 2, 3a

Psych. Abst., *27*, #499 (Rev.) A-H

Zuckerman, J. V. A note on "interest item response arrangement". *Journal of Applied Psychology*, 1953, *37*(2), 94.

Forced-Choice Items, Multiple Choice Items,
Questionnaire Theory and Development 2

ORA R-N

Zuckerman, J. V. The development of an affective adjective check list. for the measurement of anxiety. Journal of Consulting Psychology, 1960, 24, 457-462.

Adjectives, Personality Measures, Scaling	6, 14
Journal of Consulting Psychology, 24, p. 462	R-NA

465